PATENT
Docket No.: 19603/2420 (CRF D-2354)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

| | | | |
|---|---|---|---|
| Applicants | : | Stewart et al. | Examiner: M. Ibrahim |
| Serial No. | : | 09/822,080 | |
| Cnfrm. No. | : | 8901 | Art Unit: 1638 |
| Filed | : | March 30, 2001 | |
| For | : | CABBAGE PROTEINASE INHIBITOR GENE CONFERS RESISTANCE AGAINST PLANT PESTS | |



Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

**DECLARATION OF C. NEAL STEWART UNDER 37 CFR § 1.132**

Dear Sir:

I, C. NEAL STEWART, Jr., hereby declare that:

1. I received a Ph.D. degree from Virginia Polytechnic Institute and State University in December, 1993.

2. I was an Associate Professor of Biology, University of North Carolina-Greensboro, 1995-2002.

3. I am currently a Professor, Department of Plant Sciences, University of Tennessee and recipient of the Racheff Chair of Excellence in Plant Molecular Genetics.

4. I am a co-inventor of the above-identified patent application.

5. This declaration is submitted to demonstrate that the disclosure of the present application would have enabled a skilled scientist to prepare additional nucleic acid constructs having a DNA from *Brassica oleracea* encoding a Kunitz-type serine proteinase inhibitor and to use such constructs to confer insect resistance to plants by transforming plants with such constructs. Furthermore, this declaration is submitted to illustrate that the chemical and structural profiles of the Kunitz-type serine proteinase inhibitors are so well-

characterized in the literature that a skilled scientist knows what essential structural and chemical traits adhere to a protein identified as a member of the Kunitz-type serine proteinase inhibitor family.

6. Proteins that function as proteinase inhibitors are exceptional among proteins, because they tend to retain their inhibitory activity even when the active site residue ("$P_1$") is replaced by another residue (Laskowski et al., "Protein Inhibitors of Proteinases," *Ann. Rev. Biochem* 49:593-626 (1980), at 594, last para.) ("Laskowski") (attached hereto as Exhibit 1). Unlike other proteins, in which the replacement of active site residues leads to a complete loss or dramatic decrease in activity, proteinase inhibitors have the ability to tolerate a synthetic or mutational replacement of the $P_1$ residue (Id. at 607, 1st full para.). In some cases, such a substitution leads to a predictable change in inhibitory activity, e.g., an Arg63 to Trp63 substitution at the active site in the archetypal Kunitz soybean trypsin inhibitor ("STI") leads to the conversion of a trypsin inhibitor to a chymotrypsin inhibitor (Id. at 594, last para.) Thus, within each inhibitor family, the $P_1$ residue is not conserved, but changes frequently, resulting in inhibitory specificity, but not in a loss of biological activity (Id.). For example, it has been shown that generally, inhibitors with $P_1$ Lys and Arg tend to inhibit trypsin and trypsin-like enzymes, those with $P_1$ Tyr, Phe, Trp, Leu, and Met inhibit chymotrypsin and chymotrypsin-like enzymes, and those with $P_1$ Ala and Ser inhibit-elastase-like enzymes (Id. at 606, 1st full para.). Because "serine proteinase inhibitor" encompasses inhibitors of trypsin, chymotrypsin, and elastase (see Example 9), a mutation or substitution in even the active site residue of a serine proteinase inhibitor can still result in a protein with serine proteinase inhibitory activity (Laskowski at 606, 1st full para.).

7. Therefore, a nucleic acid molecule that hybridizes to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1 of my present application under stringent conditions may encode a protein with some variation in amino acid sequence from the protein encoded by SEQ ID NO: 1 while still retaining functionality as a serine proteinase inhibitor. A skilled scientist would not expect substantial variation among species encompassed by the claimed invention because of the highly stringent conditions set forth in the claims, some, but not much, variation would be expected in either the nucleotide sequence or the protein it encodes.

8. Furthermore, my patent application teaches with particularity how to make and use a nucleic acid molecule that hybridizes to the nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1 under stringent conditions.

9. Specifically, my patent application teaches the isolation of a PI gene from plant material and the preparation of a cDNA library having a PI clone (Example 1), the preparation of a shuttle vector containing the PI gene (Example 2), the transfer of the gene into a plant expression vector construct and subsequent transformation into *Agrobacterium tumefaciens* for use in transforming plant cultures (Example 3), and the characterization of the *bopi* gene, including the open reading frame of 642 bp, which encodes the mature BOPI protein (Example 4). The BoPI protein is highly characterized in the present application at Example 5. The amino acid sequence of the protein is disclosed, a putative Kunitz inhibitor family amino acid signature is identified, the Arginine active site residue at position 63 is disclosed, and the four cysteines of the expected disulfide bonds are identified (pg. 23, lines 5-28). Furthermore, the BoPI protein is characterized as having a molecular weight of 21 kDa and an isoelectric point of 4.94 (pg. 8, lines 25-29).

10. In addition, Example 6 teaches that the PI gene isolated from *Brassica oleracea* is one of a family of PI genes, thereby teaching that additional PI genes may be isolated from *Brassica oleracea*. Examples 7-10 teach how to make heterologous transgenic plants having a PI gene of the present invention, and provide methods for determining the specificity and efficacy of the proteinase inhibitor's insect antibiosis activity when a serine proteinase inhibitor is expressed in a transgenic plant.

11. Therefore, it is clear that a skilled scientist having read the present application would know: (1) how to make additional nucleic acid constructs having one or more operatively linked nucleic acid molecules which encode a Kunitz-type serine proteinase inhibitor isolated from *Brassica oleracea* having insect antibiosis activity, an operably linked heterologous DNA promoter, and an operably linked 3' regulatory region; (2) how to use such constructs to prepare expression vectors and host cells, including plant cells; and (3) how to prepare transgenic plants transformed with the construct that are resistant to insects.

12. Protease inhibitors ("PIs") from plants have been studied for over 50 years (Song et al., "Kunitz-Type Soybean Trypsin Inhibitor Revisited: Refined Structure of its Complex with Porcine Trypsin Reveals an Insight into the Interaction Between a Homologous Inhibitor From *Erythrina caffra* and a Tissue-Type Plasminogen Activator," *J. Mol. Biol.* 275:347:363 (1998) ("Song") (pg. 347 1st para. and pg. 348, 2nd full para.) (attached hereto as Exhibit 2). PIs are categorized into families, based on analogy (similar function), homology (similar amino acid structure), and mechanism of inhibiting proteinase activity (see generally, Laskowski). It is known that most individual protein inhibitors inhibit proteinases belonging to a single mechanistic class (Laskowski at 593, 1st para.). Of these,

the inhibitors of serine proteinases are the most studied (Id.) and, thus, are the most highly characterized (Id. at 601, 1st full para.). Sequencing and X-ray crystallography have shown that these inhibitors are not all homologous; rather they belong to about 10 homologous families (Id. at 601, 2nd full para.). As disclosed in my patent application, the proteinase inhibitors of the present invention belong to the Kunitz family of proteinase inhibitors (see Example 5). Kunitz-type proteins are characterized as having a molecular weight ($M_r$) of about 21,000-22,000 Daltons, two-disulfide bonds (4-half-cystine residues) and a single reactive site for serine proteases (Terada et al., "Amino Acid Sequences of Kunitz Family Subtilisin Inhibitors from Seeds of *Canavalia lineata*," *J. Biochem.* 115:397-404 (1994), first full para.) (attached hereto as Exhibit 3).

13. Another characteristic which is used to establish PI families are the topological relationships between the disulfide bonds (also known as disulfide "bridges") and the location of the PI's reactive site (Laskowski at 610, 2nd full para.). In each family, the positions of all intrachain disulfide bridges are completely conserved (Id.). Furthermore, recent X-ray crystallographic studies of the archetypal member of the Kunitz family, soybean trypsin inhibitor ("STI"), show that Kunitz-type inhibitors from a variety of plant sources share a high degree of homology to the three dimensional structure of STI (Song at pg. 347, 1st para. and pg. 348, 2nd full para.). Using the studies based on the structure of STI, it can be predicted that the reactive site of the serine proteinase inhibitors of the present invention is positioned on an exposed loop of a characteristic canonical confirmation, unconstrained by secondary structural elements or disulfide bridges that could limit its conformational freedom (Id. at 354, 1st full para.). It has been reported that in Kunitz-type mono- and dicotyledons there is clear alignment of the $P_1$ sites on the loop whether for trypsin, chymotrypsin or subtilisin (Peng et al., "Comparison of Amino Acid Sequences of the Trypsin Inhibitor From Taro (*Colocasia esculenta)*, Giant Taro *(Alocasia macrorrhiza*) and Giant Swamp Taro (*Cyrtosperma chamissonis*," *Biochem. Mol. Biol. Int'l* 31:73-81 (1993) at 80, 1st full para.) (attached hereto as Exhibit 4). In addition, proteinase inhibitors from *Brassica oleracea* have been characterized as having molecular weights ranging from 9-25 kDa, isoelectric points from 4.5-5.0, and trypsin and chymotrypsin activity that is relatively stable over a range of temperature from 0-100°C and at pH values of 4.5-7.5 (Broadway, R.M., "Purification and Partial Characterization of Trypsin/Chymotrypsin Inhibitors from Cabbage Foliage," *Phytochemistry* 33:21-27 (1993), at the Abstract ("Broadway I") (attached hereto as Exhibit 5).

14. I am familiar with W0 91/09060 to Broadway ("Broadway II") cited by the PTO. Broadway II suggests that the efficacy of a specific inhibitor from an individual plant is dependent upon 1) the unique structure of the plant proteinase inhibitor, and 2) the susceptibility of the proteinase in the target organism (see page 2, lines 19-24). While the efficacy of an inhibitor may be dependent, in part, on the inhibitor's unique structure, one can utilize an inhibitor against a susceptible target organism (i.e., a herbivorous insect) without knowledge of the inhibitor's structure because the inhibitor exhibits its functional activity against a susceptible target whether the user is cognizant of the structure of the inhibitor or not. For example, Broadway II used three PIs isolated from cabbage to successfully inhibit the growth of herbivorous insects, knowing nothing more about the proteins than their molecular weight, isoelectric point, and enzymatic inhibitory activity (see Examples I-III). It was the knowledge of the specific activity of a proteinase inhibitor that drove Broadway's choice to use a particular PI against a given insect. A skilled scientist would not consider it undue experimentation to test the specific activity of a serine PI. This can be done by carrying out an enzyme inhibition assay against trypsin, chymotrypsin, or elastase, such as described in the present application (see Example 9) or as disclosed in art (e.g., see Broadway I, starting at pg. 24, right col., second full para., to pg. 26 end of 1st partial para.), or by using a bioassay (see Example 10). Once the specificity of an isolated molecule is identified, the molecule can be used as taught in my patent application to confer insect resistance to plants. Thus, a gene, including one encoding a serine proteinase inhibitor, and the protein it encodes, can be used successfully for its intended purpose without a complete understanding of the structural character of either the nucleic acid molecule or the protein.

15. I hereby declare that all statements made herein of my own knowledge are true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that any such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Sept 24, 2003

**C Neal Stewart, Jr., Ph.D.**

*Ann. Rev. Biochem. 1980. 49:593–626*


# PROTEIN INHIBITORS OF PROTEINASES



*Michael Laskowski, Jr. and Ikunoshin Kato*

Department of Chemistry, Purdue University, West Lafayette, Indiana 47907

## CONTENTS



## PERSPECTIVES AND SUMMARY

Protein inhibitors of proteinases are ubiquitous. They are present in multiple forms in numerous tissues of animals and plants as well as in microorganisms. Their gross physiological function is the prevention of unwanted proteolysis, but detailed physiological function has been only rarely elucidated. With the exception of macroglobulins, which "inhibit" proteinases of all classes, individual protein inhibitors inhibit only proteinases belonging to a single mechanistic class. Of these the most studied are the inhibitors of serine proteinases. Most of these interact with the proteinases according to a common mechanism. In each inhibitor molecule there exists on the surface one peptide bond, the reactive site, which combines with the enzyme



0066-4154/80/0701-0593$01.00

in a substrate-like manner and serves as a substrate for the enzyme. However, while $k_{cat}/K_m$ for this interaction is large, both $k_{cat}$ and $K_m$ are, at neutral pH, very small.

In the stable enzyme-inhibitor complex the reactive site peptide bond of the inhibitor is still intact, but the carbonyl carbon is no longer trigonal. It is appreciably pyramidalized by the interaction of its oxygen with the "oxyanion hole" of the enzyme (not, as believed earlier, by the formation of a fourth bond with $O^{\gamma}$ oxygen of the catalytic seryl residue of the enzyme). The conformation of residues surrounding the reactive site peptide bond is that of an optimal substrate. Reactive sites are very rigid. Upon enzyme-inhibitor interaction conformational changes in either partner are minimal; it is almost a classical lock and key interaction.

The known inhibitors are not all homologous but consist of several families, which must have arisen by convergent rather than divergent evolution.

A striking characteristic of inhibitors is the presence of several homologous reactive sites on the same polypeptide chain, which arise from gene elongation by repeated duplication. In the case of the Bowman-Birk inhibitor family this gives rise to two homology regions, which are interconnected by interhomology region disulfide bridges. In many other cases (ovomucoids, ovoinhibitors-proteinase inhibitors from avian serum, dog submandibular inhibitor, inter-$\alpha$-trypsin inhibitors from mammalian serum) the gene doubling gives rise to several (two, three, or six) tandem domains with all disulfide bridges intradomain and only short connecting peptides connecting the domains. Chelonianin from turtle egg white consists of two tandem domains on the same polypeptide chain, but the two domains belong to two different inhibitor families.

The inhibitors are exceptional among proteins since they tend to retain their inhibitory activity upon replacement of their reactive site residue ($P_1$) by another residue. In some cases such a substitution leads to a predictable change in inhibitory activity, e.g. $Arg^{63} \rightarrow Trp^{63}$ in soybean trypsin inhibitor (Kunitz) leads to the conversion of a trypsin inhibitor to a chymotrypsin inhibitor. Such changes, which can be carried out in the laboratory, also occur during inhibitor evolution. Thus within each inhibitor family the $P_1$ residue is not conserved, but changes frequently, often resulting in changes in inhibitory specificity. In some cases (the third domains of avian ovomucoids) this residue is the most variable residue in the entire domain. This hypervariability of reactive sites during evolution stands in sharp contrast to the evolution of other proteins, where active sites are strongly conserved. However, in most other proteins, substitutions at active sites lead to the loss of activity.

# INTRODUCTION

Protein inhibitors of proteinases have not been reviewed here before. However, a book (1) and numerous reviews on this topic (2–13) have appeared elsewhere. Two international conferences on inhibitors were held (14, 15), and inhibitors were the main topic of the 23rd meeting on Protides of Biological Fluids (16). Extensive discussions of inhibitors are included in many books and symposia on the biological role of proteinases (17–22). The literature of this field is huge, with about 500 titles appearing per year. Limitations of our competence and of space led us to deemphasize in this review descriptions of partially characterized inhibitors, and of extensive studies on the physiological role and possible medical applications of protein inhibitors. We are also largely omitting the extensive and highly controversial literature on the mechanism of action of large inhibitors in mammalian blood. Because the mechanism of action of small inhibitors of serine proteinases is by now rather well known, we focus instead on the newer, extensive information on amino acid sequences of a large number of inhibitors.

While protein inhibitors of enzymes other than proteinases are well known [e.g. inhibitors of α-amylases (23), deoxyribonuclease I (24), phospholipase A (25), and protein kinases (26)], such inhibitors seem relatively rare compared to protein proteinase inhibitors. The reason for this seems easy to guess at: proteins are substrates for proteinases. All of the presently accepted and proposed mechanisms of action of protein proteinase inhibitors involve either a requirement for proteolysis or for action as a substrate analogue.

## *Limited Knowledge of Physiological Function*

The physiological function of inhibitors in the broad sense is clear—elimination of unwanted proteolysis. However, detailed functions are not clear. It is generally agreed that the presence of secretory trypsin inhibitors (Kazal) in the pancreas of vertebrates prevents premature activation of trypsinogen and, in turn, of other pancreatic zymogens. The predisposition to emphysema found in individuals with genetically determined low levels of $\alpha_1$-proteinase inhibitor strongly suggests protection against excessive proteolysis of lung tissues, but, while leukoeytic elastase is the frequently proposed target enzyme for $\alpha_1$-proteinase inhibitor, that assignment is already somewhat controversial (27). It is generally agreed that the presence of large numbers of proteinase inhibitors in massive amounts in mammalian plasma serves to delimit blood clotting (28) and other proteolytic cascade

processes, such as hormone production and elimination and complement fixation (29). However, the precise physiological involvement of inhibitors in these processes is not known.

A good case has been made that many proteinase inhibitors in plants serve a defensive function against insect infestation by inhibition of insect proteinases. Ryan demonstrated that wounding of potato leaves leads to great increases in the level of a polysaccharide hormone, proteinase inhibitor-inducing factor (PIIF), which in turn leads to huge increases in inhibitor levels (30, 31). Many other exciting studies on the role of inhibitors are reported but in most cases the results, while suggestive, are far from unequivocal. In some cases, added research complicates rather than supports the naive original assignments. Thus an important role was at first envisaged for acrosin inhibitors in the process of sperm capacitation, but more recent work casts doubt on this conclusion (32).

For other inhibitors the situation is far worse, mainly because proteinase inhibitors are most frequently discovered and isolated by noting their ability to inhibit some readily available proteinase, e.g. bovine trypsin, rather than physiological target enzyme. They thus become proteins in search of a function, rather than proteins isolated to account for a previously discovered biological function. Because most laboratories used only bovine trypsin as the initial test enzyme, an opinion formed that most proteinase inhibitors are inhibitors of trypsin. As chymotrypsin, elastase, and subtilisin were added as screening tools, many inhibitors of these enzymes were discovered. None of these are likely to be true target enzymes for most of the isolated inhibitors. The lack of knowledge of true target enzymes for the majority of inhibitors is now one of the major stumbling blocks in the understanding of inhibitor evolution and of inhibitory specificity. It appears to us quite likely that the target enzyme problem can be best solved for intracellular proteinases and for their cognate inhibitors (33, 34).

## *Nomenclature Problems*

In a review published nine years ago we deplored the confusion in inhibitor nomenclature (2). Since then a great deal about inhibitors has been learned, a huge number of new inhibitors discovered, and the confusion grows worse. Newly discovered inhibitors are usually named according to the biological source and the inhibited enzyme [e.g. S-SI, *Streptomyces* subtilisin inhibitor (35); plasminostreptin, plasmin inhibitor from *Streptomyces* (36)]. Both procedures cause confusion: the biological source because inhibitors belonging to more than one family may be present in the same source. In such cases we have suggested the use of the last name of the initial

discoverer as a distinguishing method (e.g. soybean trypsin inhibitor (Kunitz) and soybean trypsin inhibitor (Bowman-Birk); pancreatic trypsin inhibitor (Kunitz) and pancreatic secretory inhibitor (Kazal)]. However, this is not sufficient. In soybeans there exist many inhibitors of the Bowman-Birk family and these are now simply designated by letters and numbers (37). The Kunitz inhibitor from pancreas is not organ-specific but occurs in many bovine organs and most recently has been assigned to mast cells (38). The recent finding that the inhibitor in chicken plasma and chicken ovoinhibitor probably have the same amino acid sequence poses a serious problem (39, 40). The two inhibitors should probably be assigned the same name, but the well-established name, ovoinhibitor (41), seems grossly inappropriate for a plasma inhibitor. Naming after biological source clearly produces confusion and occasionally leads to the wasted effort of characterizing the same inhibitor several times. Even greater problems stem from the use of the inhibited enzyme as a part of the name. First, the discoverer usually does not know the true target enzyme or the enzyme that is most strongly inhibited. Thus $\alpha_1$-proteinase inhibitor was first named $\alpha_1$-antitrypsin, when in fact many workers suspect that trypsin is not an important target enzyme for this inhibitor nor, among the many enzymes it inhibits, is trypsin most rapidly or strongly inhibited (42). Yet, because of the name, most studies of enzyme interactions with this inhibitor focus on trypsin. The suggestion to change the name to $\alpha_1$-proteinase inhibitor seems to have only limited acceptance. Another objection to specifying the inhibited enzyme as part of the name of the inhibitor is the recently discovered phenomenon of reactive site variability (37, 43–45). Frequently, highly homologous inhibitors from closely related species inhibit different enzymes. When this is the case, a search for a homologue may well fail if one is looking, for example, for trypsin inhibition, and the closely related homologue is now an elastase inhibitor (e.g. chicken and bobwhite quail ovomucoids).

Another strategy, which seemed superior, was to assign to inhibitors totally trivial names such as Trasylol®, Bayer's trademark for pancreatic trypsin inhibitor (Kunitz), or ones that indicate only the source and inhibition but do not specify the enzyme, such as ovomucoid and ovoinhibitor.

We suggest that a rational system of nomenclature should involve first assigning the inhibitor to a family, then specifying the number of domains (if more than one) and the species from which it is isolated. The additional designators, such as its tissue source and the enzyme inhibited should be added last, since they are most likely to be confusing. However, as the list of families is still quite preliminary and since assignment to a family requires sequencing, introduction of the new nomenclature is still premature.

## $\alpha_2$ MACROGLOBULINS

The term protein inhibitor of proteinases could in principle denote any protein that diminishes the enzymatic activity of a proteinase. The most bizarre examples, such as proteinases that "inhibit" other proteinases by digesting them, fortunately are not called proteinase inhibitors. However, two entirely dissimilar groups of proteins are called proteinase inhibitors. One of them consists of the macroglobulins, high-molecular-weight proteins apparently present in the plasma of all mammals, often in multiple forms. They are exemplified by $\alpha_2$ human macroglobulin, often abbreviated as $\alpha_2$M, which has a molecular weight of 720,000 and consists of four apparently identical polypeptide chains. It dissociates into pairs of chains upon denaturation and into single chains upon reduction of disulfide bridges (46, 47).

When $\alpha_2$M combines with proteinases, only the proteolytic activity toward large protein substrates is decreased or eliminated. The activity toward specific burst substrates and small synthetic substrates is unimpeded (48). Strikingly, the $\alpha_2$M-proteinase complexes can still be inhibited by small protein proteinase inhibitors, such as pancreatic trypsin inhibitor (Kunitz) or pancreatic secretory trypsin inhibitor (Kazal), but not by larger ones such as soybean trypsin inhibitor (Kunitz) (49). Clearly, in the $\alpha_2$M complexes, the active site of the proteinase is quite open; it is the access of substrates and inhibitors to this site that is sterically hindered. The other remarkable feature of macroglobulins is their ability to bind a wide variety of proteinases belonging to all four mechanistic classes (50). Indeed, it is relatively rare to find endoproteinases that do not form complexes with macroglobulins. The above facts gave rise to the "trap" mechanism of macroglobulin action (51). The proteinase hydrolyzes one or more particularly susceptible peptide bonds in $\alpha_2$M, and triggers a conformational change in $\alpha_2$M which traps the enzyme molecule. Only molecules of very high molecular weight can act by this mechanism. To account for the broad specificity the critical peptide segment must be rather long since it must contain peptide bonds matching the specificities of the various proteinases. Evidence has been provided for specific proteolytic cleavages upon complex formation.

Much research in vivo and in vitro compares the importance of $\alpha_2$M with that of other inhibitors present in blood. Upon addition of a small quantity of enzyme to a mixture of $\alpha_2$M and another inhibitor, complexes with both inhibitors are formed in a ratio that depends upon the relative values of the second order rate constants for association. However, the complex with a typical inhibitor dissociates, albeit slowly, while the $\alpha_2$M complex is "irre-

versible" and does not. The result in an in vitro experiment is that all of the proteinase is ultimately transferred to an $\alpha_2$M complex, unless there is so much enzyme present that $\alpha_2$M is saturated. In vivo this effect is greatly exaggerated, since $\alpha_2$M proteinase complexes are cleared very rapidly; the half-life is about 10 min in man (52). Thus, essentially all injected proteinases are cleared by the $\alpha_2$M pathway. Other serine proteinase inhibitors are only transiently involved in transferring the proteinases to be cleared. There is considerable interest in understanding the signals leading to the more rapid clearance of complexed as opposed to free $\alpha_2$M. The view presented above is an idealization and probably does not apply to all serine proteinases.

The research on macroglobulins clearly shows that they are of central importance in processing of proteinases in blood. However, their interaction with proteinases is so different from that of other protein proteinase inhibitors that at least to chemists their designation as inhibitors seems undesirable.

## INHIBITORS WITH CLASS-SPECIFIC REACTIVE SITES

The remaining protein proteinase inhibitors differ strikingly from the macroglobulins. For an overwhelming majority of them the following three statements are justified:

1. In the enzyme inhibitor complexes all enzymatic activities toward all substrates are totally abolished.
2. The inhibition is strictly competitive.
3. A particular inhibitory reactive site can inhibit only proteinases belonging to a single one of the four mechanistic classes proposed by Hartley (50).

All of these statements are frequently contradicted in the literature, mainly because of experimental errors, in our opinion (see 2). We are not aware of any well-characterized system that contradicts any of these three statements. If point 3 is granted then it is most convenient to divide the inhibitors into four major classes.

### *Inhibitors of Carboxyl, Metallo and Sulfhydryl Proteinases*

Many inhibitors of these proteinases have been reported, some of which are listed in Table 1. However, they remain an understudied group. Complete amino acid sequences are known for only two such inhibitors; three-dimensional structures for none. In spite of a few elegant studies there is no definitive proposal for the mechanism of action of any of the inhibitors listed in Table 1.

Table 1 Some inhibitors of thiol, carboxyl, and metallo-proteinases

| Enzyme inhibited | Source of inhibitor | Molecular weight | Ref |
|---|---|---|---|
| **Thiol proteinase** | | | |
| Ficin, papain | Avian egg white | 12,700 | 53, 54 |
| Cathepsin B1, C | Avian egg white | | 55, 56 |
| Bromelain | Pineapple stem | 5,600 | 57 |
| Papain, ficin, bromelain | Rat skin | 74,000 | 58 |
| | | 13,400 | |
| Papain | Rabbit skin | 12,500 | 59, 60 |
| Papain, ficin, bromelain | *Bauhinia* seeds | 24,000 | 61 |
| Papain, cathepsin B1 | Leucocyte cell, spleen cell | 15,000 | 62 |
| Papain, cathepsin B1 | Bovine nasal cartilage | 13,000 | 63 |
| Ficin, papain | $\alpha_2$ Serum protein (human) | 90,000 | 64, 65 |
| Human liver cathepsin B1 | Immunoglobulin G | | 66 |
| **Carboxyl proteinase** | | | |
| Pepsin | *Ascaris lumbriocoides* | 15,500–17,500 | 67–69 |
| Cathepsin D, E | *Ascaris lumbriocoides* | | 70 |
| Cathepsin D | Potato | 27,000 | 71, 72 |
| Pepsin | *Bauhinia* seeds | 24,000 | 61 |
| Pepsin | *Scopolia japonica* cultured cell | 4,000–6,000 | 73 |
| **Metallo proteinase** | | | |
| Collagenase (human) | Cartilage, aorta (bovine) | 11,000 | 74, 75 |
| | Bovine nasal cartilage | 22,000 | 63 |
| | Rabbit tissues in culture | 27,500–29,500 | 76 |
| | Porcine aorta smooth muscle cell in culture | | 77 |
| Collagenase (human gastric mucosal) | $\beta_1$ serum protein | 40,000 | 78 |
| $Ca^{2+}$-dependent neutral proteinase | Rat liver | 30,000 | 79 |
| Neutral metallo proteinase | Rabbit bone | | 80 |
| Carboxypeptidase A, B | Potato | 4,300 | 81–85 |
| Carboxypeptidase A | *Ascaris lumbriocoides* | 7,530 | 67, 86 |
| Aminopeptidase | *N. crassa* | 10,000 | 87, 88 |
| | | 5,000 | |

## *Inhibitors of Serine Proteinases*

The number of well-characterized inhibitors of serine proteinases far exceeds the number of described inhibitors of the three other mechanistic classes of proteinases. It is not clear whether this relative abundance is a true reflection of distribution of inhibitors in nature or only of the preferences and convenience of the biochemists who isolate them.

In contrast to other inhibitors, the mechanism of interaction of protein inhibitors of serine proteinases with their cognate enzymes is known in considerable detail. Sequencing and X-ray crystallographic studies have shown that these inhibitors are not all homologous; instead they consist of several (probably at least 10) homologous families. Surprisingly, however, most inhibitors of serine proteinases ($\alpha_2$ macroglobulin excluded) interact according to the "standard mechanism" described below. This situation is analogous to the existence of at least two nonhomologous families of serine proteinases, the subtilisin and the chymotrypsin families, which do not share a common three-dimensional structure, but which hydrolyze their substrates and are inhibited by inhibitors according to the same mechanism. It is not yet possible to decide whether several heavily studied inhibitors—the rather small stable thrombin inhibitor from leeches, hirudin (89), and the class of large, relatively unstable inhibitors from mammalian plasma—also act by the standard mechanism. These are examined briefly at the end of this review.

## *The Standard Mechanism*

Inhibitors obeying this mechanism are highly specific, limited proteolysis substrates for their target enzymes. On the surface of each inhibitor molecule lies at least one (more in multiheaded inhibitors) peptide bond called the reactive site (90), which specifically interacts with the active site of the cognate enzyme. The value of $k_{cat}/K_m$ for the hydrolysis of this peptide bond by the cognate enzyme at neutral pH is very high, $10^4$–$10^6$ $M^{-1}$ $s^{-1}$ (91, 92), compared to a typical value for normal substrates of about $10^3$ $M^{-1}$ $s^{-1}$. However, for inhibitors, the values of $k_{cat}$ and $K_m$ are both many orders of magnitude lower than the values for normal substrates. At typically used concentrations and neutral pH, therefore, their hydrolysis is extremely slow, and the system behaves as if it were a simple equilibrium between the enzyme and free inhibitor on the one hand and the complex on the other. The equilibrium constant for the association is extremely high (in the range of $10^7$–$10^{13}$ $M^{-1}$) (91, 2). An additional property of the inhibitory reactive sites is that their hydrolysis does not proceed to virtual completion. Instead, a neutral pH, the equilibrium constant between modified inhibitor (reactive site peptide bond hydrolyzed) and virgin inhibitor (reactive site peptide

bond intact) is near unity (92, 93–98). Since the same stable complex is formed between the enzyme and either modified or virgin inhibitor, both are thermodynamically equally strong inhibitors of the cognate enzyme (2). However, in cases so far examined the rate of complex formation from modified inhibitor and the enzyme is much lower than from virgin inhibitor and the enzyme. In a few cases, this difference is so great that it led to the assumption that the modified inhibitor was inactive, which longer incubation showed to be an error.

That the specifically hydrolyzed peptide bond in the inhibitor is the reactive site is also shown by the following experiments: (*a*) While in every case tried the modified inhibitor is active, specific removal of its newly formed carboxyl terminal residue ($P_1$) always renders it inactive (99). (*b*) Specific acylation of the newly formed amino terminal residue ($P_1'$) prevents return to the virgin inhibitor and thus renders the inhibitor inactive (100). (*c*) Kinetically controlled dissociation of the enzyme-inhibitor complex prepared with modified inhibitor leads to the enzyme and a mixture of virgin (predominant) and modified (minor) inhibitor (101–103).

The overall mechanism of the enzyme-inhibitor interaction, including only those intermediates whose existence was definitely shown, can be written as (91, 99, 101, 103, 110, 111):

$$E + I \rightleftharpoons L \rightleftharpoons C \rightleftharpoons X \rightleftharpoons L^* \rightleftharpoons E + I^*$$

where E is the enzyme, I and I* virgin and modified inhibitors, respectively, L and L* are loose, noncovalent (rapidly dissociable) complexes (110) of E with I and I*, respectively, X is the relatively long-lived intermediate in the E + I*, reaction (111), and C is the stable enzyme-inhibitor complex. It is clear that more intermediates will be found as finer and finer techniques are applied to the analysis of this mechanism.

Of greatest interest to most investigators is the detailed nature of the stable complex, C. While a number of inferences can be made from thermodynamic, kinetic, and chemical modification data, the most penetrating analyses come from X-ray crystallographic examination of complexes. Three enzyme inhibitor systems have now been analyzed by X-ray crystallography: soybean trypsin inhibitor (Kunitz) with porcine trypsin (112); *Streptomyces* subtilisin inhibitor (113), S-SI, with subtilisin (114); and pancreatic trypsin inhibitor (Kunitz) (115, 116) with bovine trypsin (117, 118), bovine anhydrotrypsin (119), and bovine trypsinogen in the presence and absence of isoleucylvaline (120). All of these studies show that the reactive site of the inhibitor (as previously identified by chemical studies) reacts with the active site of the enzyme in a substrate-like manner. In each case contact occurs over a small portion of the enzyme and the inhibitor, but over this

portion the fit is excellent and numerous van der Waals interactions, hydrogen bonds, and salt bridges are formed. In each case complex formation occurs with relatively small conformational change; it is predominantly a lock and key interaction, although a small conformational adaptation of the inhibitor (presumably in the L → C step) does take place.

The detailed molecular events at the active site of the enzyme and at the reactive site of the inhibitor are not as clear as some facile statements imply. Here, only the highest resolution work of the group led by Huber, summarized in several reviews (121–123), is considered. The scissile peptide bond ($P_1$-$P_1'$) of the inhibitor is intact in the complex. However, the carbonyl carbon of the reactive site peptide bond ($P_1$ residue) is not fully trigonal but is distorted about half way toward tetrahedral. (The older view that it was fully tetrahedral disappeared with refinement.) Since the carbon is 2.6 Å (much longer than the 1.4 Å covalent C–O bond, but shorter than a van der Waals contact) from the $O^{\gamma}$ oxygen in the catalytic $Ser^{195}$ of the enzyme, it was believed that it was this oxygen that caused the tetrahedral distortion of the carbonyl carbon. This view was abandoned as it was realized that anhydrotrypsin (124) (where the catalytic $Ser^{195}$ residue is dehydrated to dehydroalanyl$^{195}$) forms stable inhibitor complexes of comparable strength (124), with essentially the same geometry (119), and with tetrahedral distortion precisely the same as in a complex with trypsin (118). Therefore the partial tetrahedral distortion is currently believed to result from the attraction between the carbonyl oxygen of the inhibitor and the oxyanion hole (125) of the enzyme, the NH's of $Gly^{193}$ and $Ser^{195}$. The carbonyl carbon is now poised for nucleophilic attack by oxygen $O^{\gamma}$ of $Ser^{195}$, but in the stable complex the attack has not yet taken place.

In recent studies $^{13}C$ was introduced into the carbonyl carbon of the reactive site (1-carbon of the $P_1$ residue) of soybean trypsin inhibitor (Kunitz) (126, 127) and of pancreatic trypsin inhibitor (Kunitz) (R. Richarz, H. Tschesche, and K. Wuthrich, personal communication). If the carbonyl carbon changed from trigonal in the free inhibitor to the completely covalent tetrahedral intermediate in the enzyme inhibitor complex (the oldest of the three models) complex formation should be accompanied by a huge (~50 ppm) upfield chemical shift. In fact, a very small (~1 ppm) downfield chemical shift is observed by all three groups that did the experiment. The complete tetrahedral intermediate is thus excluded. It is not yet clear whether these results are compatible with the X-ray results as currently interpreted.

In our opinion the X-ray crystallographic studies do not explain as yet why inhibitors are inhibitors, not just excellent substrates. This is a complex problem, since strong inhibitors of bovine trypsin are simply very good substrates for a homologous trypsin 1 from the starfish *Dermasterias im-*

*bricata* (92, 135). Human cationic trypsin is intermediate in behavior between these two enzymes (129). Thus, the high $k_{cat}/K_m$ value of the reactive site of an inhibitor applies to many homologous enzymes. Some of these partition this value into very low $k_{cat}$ and $K_m$, and thus are strongly inhibited; others show more conventional $k_{cat}$ and $K_m$ values and the inhibitor serves them as a substrate.

THE NATURE OF THE REACTIVE SITE The X-ray crystallographic studies of pancreatic trypsin inhibitor (Kunitz) (116), soybean trypsin inhibitor (Kunitz) (112), and subtilisin inhibitor, S-SI (113), permit comparison of the three-dimensional structure of these inhibitors. The overall folding of all three is dramatically different (114); they thus belong to three separate families, strongly supporting the notion that the inhibitor families arose by convergent evolution. In contrast, as seen in Table 2, the geometry of the region surrounding the reactive site (114) is virtually identical for the three inhibitors even though the sequences are quite different, and is also very similar to the "hypothetical subtilisin substrate" (128). As the experimental errors in some of these figures are large, the similarities rather than differences deserve stress at this time. The results support the notions that (*a*) inhibitors are predominantly very good substrates and (*b*) the mechanism of interaction is closely similar for enzymes belonging to the subtilisin and chymotrypsin families, and for inhibitors belonging to the various inhibitor families. The amino acid sequences surrounding the reactive sites of the various families are summarized in Table 3.

In all inhibitors that clearly obey the standard mechanism, the reactive site peptide bond is encompassed in at least one disulfide loop (90, 2), which ensures that during conversion of virgin to modified inhibitor the two peptide chains cannot dissociate. However, this requirement is not absolute since soybean trypsin inhibitor (Kunitz) with its $Met^{64}$-Leu peptide bond hydrolyzed by subtilisin is still fully inhibitory, and can be converted to

Table 2 Reactive site residues and main-chain conformations of BPTI, STI, and S-SI (114)

| | $P_3$ | | $P_2$ | | $P_1$ | | $P_1'$ | | $P_2'$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\phi$ | $\psi$ | $\phi$ | $\psi$ | $\phi$ | $\psi$ | $\phi$ | $\psi$ | $\phi$ | $\psi$ |
| BPTI | −86 | −28 | −64 | 152 | −116 | 87 | −135 | 166 | −116 | 87 |
| STI | −46 | −21 | −77 | 136 | −89 | 85 | (−135) | (−115) | −175 | −179 |
| S-SI | −123 | 141 | −52 | 134 | −92 | 89 | −117 | 169 | −120 | 78 |
| Hypothetical substrate for subtilisin | (−157) | (172) | (−84) | (162) | (−116) | (26) | (−62) | (136) | (−112) | (−70) |

Table 3 Alternative amino acid residues in the sequences surrounding the reactive site (arrow) of inhibitors

| | $P_4$ | $P_3$ | $P_2$ | $P_1$ | $P'_1$ | $P'_2$ | $P'_3$ | $P'_4$ | $P'_5$ | $P'_6$ | $P'_7$ | $P'_8$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pancreatic Trypsin Inhibitor (Kunitz) Family[a] | ...G | P | C | K | A | R | I | I | R | Y | F | Y... |
| | | L | | R | G | A | L | P | Q | F | Y | F |
| | | R | | Y | R | S | F | L | S | W | A | |
| | | N | | M | Q | Y | T | R | A | I | H | |
| | | | | L | | K | V | Q | L | | | |
| | | | | - | | F | | S | | | | |
| | | | | | | H | | | | | | |
| | | | | | | M | | | | | | |
| | | | | | | P | | | | | | |
| Pancreatic Secretory Trypsin Inhibitor (Kazal) Family[b] | ...G | C | P | R | I | Y | N | P | V | C... | | |
| | V | | N | K | D | L | R | L | I | | | |
| | A | | T | E | E | P | H | F | L | | | |
| | L | | A | D | A | F | Q | R | H | | | |
| | D | | M | A | L | H | S | E | | | | |
| | M | | L | L | Q | D | M | | | | | |
| | F | | S | S | N | Q | K | | | | | |
| | | | | M | | M | D | | | | | |
| | | | | V | | | I | | | | | |
| | | | | Q | | | E | | | | | |
| | | | | Y | | | | | | | | |
| Streptomyces Subtilisin Inhibitor Family[c] | ...M | C | P | M | V | Y | D | P | V | L... | | |
| | A | | T | K | Q | F | | | | V | | |
| Bowman-Birk Inhibitor Family[d] | ...A | C | T | K | S | N | P | P | Q | C... | | |
| | M | | A | R | | M | | G | K | | | |
| | V | | | A | | I | | A | T | | | |
| | L | | | L | | Q | | | | | | |
| | I | | | F | | Y | | | | | | |
| | S | | | Y | | | | | | | | |
| Soybean Trypsin Inhibitor (Kunitz) Family[e] | ...P | S | Y | R | I | R | F | I | A | E... | | |

Table 3 *(Continued)*

| | $P_4$ | $P_3$ | $P_2$ | $P_1$ | $P_1'$ | $P_2'$ | $P_3'$ | $P_4'$ | $P_5'$ | $P_6'$ | $P_7'$ | $P_8'$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Potato Inhibitor I Family[f] | ...P | V | T | L | D | Y | R | C | N | R... | | |
| | | | | M | | F | | | | | | |
| | | | | | | L | | | | | | |
| Potato Inhibitor II Family[f] | ...A | S | Y | K | S | V | C | E | G | E... | | |
| | | - | | | | - | | | | | | |
| Alpha-1-Proteinase[g] Inhibitor Family | ...A | I | P | M | T | I | P | P | E | V... | | |
| | | | | | S | | | | | | | |

[a]Inhibitors of Kunitz family were obtained from: bovine pancreas (148); red sea turtle egg white, chelonianin (149); bovine colostrum (150); *Helix pomatia* K (151); venom of *Vipera russelli* (152); venom of *Hemachatus haemachatus* (153); venom of *Naja nivea* (153); various toxins from the venom of *Dendroaspis polylepis polylepis* toxin E (154); toxin K (155), and toxin I (156); sea anemone, *Anemonia sulcata* (157); $\beta$ chain of $\beta_1$-bungarotoxin (146); and bovine and human inter-$\alpha$-trypsin inhibitor (158, 159).

[b]Inhibitors of Kazal family were obtained from: bovine pancreas (138); ovine pancreas (160); porcine pancreas (161); canine pancreas (K. Hochstrasser, unpublished); human pancreas (162); turkey pancreas (163); guinea pig seminal plasma (166); leech (*Hirudo medicinalis*) (141); canine submandibular glands (164, 165); ovomucoids (avian egg white) (44); Japanese quail and chicken ovoinhibitor (egg white) and chicken $\alpha_2$-proteinase inhibitor (serum) (39); and porcine seminal plasma (166).

[c]Inhibitors of *Streptomyces* subtilisin inhibitor family were obtained from: *Streptomyces albogriseolus* (137) and *Streptomyces antifibrinolyticus* (136).

[d]Inhibitors of Bowman-Birk inhibitor family were obtained from: lima bean IV, IV' (168, 169); soybean Bowman-Birk inhibitor (170); soybean D–II, E–I (171); soybean C–II (172); garden bean II and II' (173); chick pea (174); adzuki bean I (175); and *Vicia angustifolia* (176).

[e]Inhibitor of soybean trysin inhibitor (Kunitz) family was obtained from soybeans (167).

[f]Inhibitors of potato inhibitors I and II families were obtained from: var. Russet Burbank (177, 178); var. Dan Shaku Imo (179–181).

[g]Inhibitor of $\alpha_1$-proteinase inhibitor family was obtained from human plasma (182, 183).

modified inhibitor ($Arg^{63}$-Ile reactive site hydrolyzed) without loss of the 64–84 fragment (129). Clearly, strong noncovalent bonding can substitute for a disulfide bridge. Thus, eglin (130) and $\alpha_1$ proteinase inhibitor (131), which are devoid of intramolecular disulfide bridges, could still obey the standard mechanism. Alternatively, the standard mechanism may have to be modified to allow for release of an inhibitor fragment from complexes formed with these inhibitors.

The reactive site residue, $P_1$, generally corresponds to the specificity of the cognate enzyme. Thus inhibitors with $P_1$ Lys and Arg tend to inhibit trypsin and trypsin-like enzymes (2), those with $P_1$ Tyr, Phe, Trp (artificial only), Leu and Met inhibit chymotrypsin and chymotrypsin-like enzymes, and those with $P_1$ Ala and Ser inhibit elastase-like enzymes. Those that inhibit chymotrypsin often, but not always, inhibit subtilisin. These, however, are

very rough rules. In the Kazal secretory inhibitor family, inhibitors with $P_1$ Leu and Met are strong inhibitors of elastase (44); in the Bowman-Birk family elastase is inhibited with $P_1$ Ala, but not with $P_1$ Leu (37, 132). In many instances strong inhibitors of trypsin with $P_1$ Arg and Lys inhibit chymotrypsin on the same reactive site (37, 133, 134). In other cases related inhibitors do not inhibit chymotrypsin, and failure to inhibit is not yet fully predictable. The most striking feature of $P_1$ specificity is that exchange of Lys for Arg at this position, either by actual mutation or by semisynthetic replacement (104, 107), leaves the inhibitor specificity and strength approximately the same. The exchange of Lys or Arg for a chymotrypsin specific residue generally changes the inhibitor from a good trypsin inhibitor to a good chymotrypsin inhibitor (105, 107). An odd exception is the rather strange behavior of $Phe^{63}$ soybean trypsin inhibitor (Kunitz) (105, 106).

The ability to tolerate a synthetic or mutational replacement of the $P_1$ residue and still retain inhibitory activity, either with retention of original specificity (e.g. after Arg → Lys or Leu → Met exchanges) or with dramatic change in specificity (e.g. tryptic to chymotryptic after Arg ⇆ Trp exchanges), is peculiar to proteinase inhibitors. In most biologically active proteins the replacement of active site residues, even by closely related ones, leads to a complete loss or dramatic decrease in activity. We believe that largely because of this chemical difference the evolutionary behavior of most proteins and inhibitors differ. In most proteins, active site residues are stubbornly conserved, whereas in every inhibitor family when enough examples are known, there is appreciable variation in $P_1$ residues. This occurs often even between orthologous proteins from closely related species [e.g. plasminostreptin (Lys) (136) and S-SI (Met) (137); bovine (Arg) (138) and porcine (139) (Lys) pancreatic secretory inhibitors (Kazal), etc]. The phenomenon takes its extreme form in third domains of avian ovomucoids (44), where the $P_1$ residue is the most subject to evolutionary change of all 56 residues in the domain—a phenomenon referred to as hypervariability of the reactive site. This case is only an extreme example of behavior observed in the evolution of all inhibitors.

Although contrary to our earlier opinion, the $P_1'$ residue in most inhibitor families can tolerate a broad range of residues, there are two exceptions: (*a*) It is generally agreed that serine proteinases seldom hydrolyze bonds with $P_1'$ Pro. Apparently, inhibitors with $P_1'$ Pro are not active. Silver pheasant ovomucoid with ... Cys Asn Lys ↓ Ala ... reactive site sequence inhibits trypsin; highly homologous golden pheasant ovomucoid with ... Cys Asn Lys ↓ Pro ... does not (W. J. Kohr and M. Laskowski, Jr., unpublished). Except for this case $P_1'$ Pro is not observed. (*b*) There is stubborn conservation of $P_1'$ Ser in all known Bowman-Birk inhibitors. Semisynthetic substitution study (109) showed that Ser in this position was

better than any other amino acid residue whose substitution was attempted. Rather surprisingly a synthetic variant of pancreatic trypsin inhibitor (Kunitz) with ... Cys Lys ↓ Phe ... reactive site (140) is a much weaker inhibitor than the natural ... Cys Lys ↓ Ala ... inhibitor.

Generalizations about other positions near the reactive sites are harder to draw. In all active inhibitors (note that the tables include reactive site sequences of some inactive molecules clearly homologous to active inhibitors) the residues in positions $P_2$ are neutral, consistent with their burial in the enzyme-inhibitor interface. Where $P_2$ is not required to be Cys, it is quite often Thr or a single base mutant of this amino acid. Contrary to our earlier claim, Pro is not required in the neighborhood of the reactive site as bdellin (141), a trypsin inhibitor from leeches, has no Pro residues. However, apparently at several positions ($P_3$, $P_2$, and probably $P_3'$ and $P_4'$) Pro helps to ensure the proper reactive site geometry, even though this can be obtained without Pro (see Tables 2 and 3).

Semisynthetic modifications also led us to the conclusion (108, 142) that in soybean trypsin inhibitor (Kunitz) removal of the amino acid residue $P_1'$ leads to inactivation of the inhibitor, presumably because reformation of the $P_1$–$P_2'$ peptide bond is not possible. Similarly, insertion of an additional residue between $P_1$ and $P_1'$ also led to inactivation even though all the peptide bonds were reformed, presumably because the required reactive site geometry (Table 2) was destroyed.

## MULTIPLE REACTIVE SITES ON A SINGLE POLYPEPTIDE CHAIN

Where two (or more) enzymes can be inhibited by the same inhibitor molecule, two operationally distinguishable situations can arise. The enzymes can compete for the inhibitor (such inhibitors are unfortunately referred to as polyvalent) or they can be inhibited simultaneously and independently. Both situations are quite common. In the competitive case we can envisage either inhibition on the same reactive site or inhibition on two distinct, overlapping reactive sites where steric hindrance prevents association with both enzymes. Overlapping reactive sites seem quite plausible, but years of searching have yielded no convincing examples of such inhibitors of known amino acid sequence. However, such an explanation continues to be attractive for many less well-known inhibitors.

Numerous cases of independent inhibition of several enzymes are known. Functionally, inhibitors are classified as single-headed if they have only one active reactive site, double-headed if they have two and so on (143). Multiheadedness results from the presence of several active sites on the same molecule, and is most commonly achieved in protein chemistry by noncovalent association of several peptide chains each with one active site. Multiheadedness is achieved this way for at least two inhibitors. The subtilisin

inhibitor, S-SI, is a strong noncovalent dimer (144) that inhibits two molecules of subtilisin without dissociation. Similarly, potato inhibitor I is a tetramer which inhibits four molecules of chymotrypsin (145). While several other inhibitors associate (most notably Bowman-Birk inhibitors), they apparently dissociate during complex formation. Multiheadedness also results from the presence of two separate chains linked by one or more disulfide bridges. Turtle egg white inhibitor testudin [a Kazal inhibitor (I. Kato and M. Laskowski, Jr., unpublished results)] and $\beta_1$ bungarotoxin (146) (a Kunitz PTI type inactive inhibitor) are partial examples of such inhibitors.

The most common method of achieving multiheadedness is by gene duplication. The first such case discovered was lima bean inhibitor (147), a member of the Bowman-Birk inhibitor family. Inhibitors of this family consist of two homology regions, each with one reactive site in a homologous position. Even more striking examples are found in the Kunitz and Kazal families of inhibitors. In the Kunitz family, the COOH terminus of the inter-$\alpha$-trypsin inhibitor consists of two tandem domains connected only by a short connecting peptide. Dog submandibular inhibitor consists of two, all avian ovomucoids of three, and avian ovoinhibitors (also inhibitors from avian serum) of six tandem Kazal type domains. There is weaker evidence for several other inhibitors with multiple tandem homologous domains. In most cases, domains can be separated by cleavage of a single peptide bond in the connecting peptide and are independently active.

The tendency to form multiheaded inhibitors by gene elongation through partial or complete duplication (often repeated several times) is a general tendency of protein protinase inhibitors and stands in reasonably sharp contrast with most biologically active proteins where a single active site in a single polypeptide chain is most common. The minimal chemical requirement for such behavior is the dispensability of the $NH_3^+$ terminal group and of the $\overline{COO}^-$ terminus for activity. However, the biological driving force is, in many cases, obscure. It is clear for the plasma inhibitors—the inter-$\alpha$-trypsin inhibitor (158, 159) in mammalian plasma and the six Kazal domain inhibitor (39) in chicken plasma—that formation of multiple domains raises the molecular weight and prevents excretion and yet does not require the use of a large number of amino acids per reactive site. In fact, it is surprising that several mammalian plasma inhibitors (e.g. $\alpha_1$ proteinase inhibitor, $\alpha_1$ antichymotrypsin) do not have multiple reactive sites. No similar rationalizations are available for the Bowman-Birk inhibitors, for ovomucoids, and for dog submandibular inhibitor where the physiological functions are unknown.

Gene elongation by duplication is not the only way in which inhibitors become multiheaded. Red sea turtle inhibitor, chelonianin (Kunitz type),

(149) consists of two domains belonging to two different inhibitor families. Thus, elongation by gene fusion is another possible mechanism. The extent to which this mechanism is used is not yet known.

## *The Inhibitor Families*

The number of known and partially characterized inhibitors of serine proteinases is enormous. They can be classified into at least 10 and probably more families. To conserve space we adopt the conservative view and describe only families for which at least one sequence has been unambiguously determined, where the reactive site has been cleanly assigned, and where there is clear evidence that the standard mechanism is obeyed. The tentative list of such families is presented in Table 4.

While the major criterion for establishing a family is extensive homology among its members, the easiest family characteristics to appreciate at a glance are the topological relationships between the disulfide bridges and the location of the reactive site. These are presented in Figure 1. In each family (except possibly potato II and *Ascaris* trypsin inhibitor) the positions of all intrachain disulfide bridges are completely conserved. A few inhibitors (or inhibitor homologues) have additional interchain disulfide bridges (e.g. testudin and $\beta_1$ bungarotoxin).

THE PANCREATIC TRYPSIN INHIBITOR (KUNITZ) FAMILY This family is named after the first inhibitor to be isolated in crystalline form (184), the first for which the typical 1 : 1 enzyme-inhibitor stoichiometry was determined (184, 185), the first for which reversibility was demonstrated (184), the first to be sequenced (148), and the first whose three-dimensional structure was determined crystallographically both for the free inhibitor and for the enzyme-inhibitor complex (115–119). This was the first, and to date the only inhibitor to be synthesized in the laboratory (140, 186, 187). Since it is small (58 amino acid residues), soluble, stable, and its sequence

Table 4 Families of protein inhibitors that inhibit serine proteinases and obey the standard mechanism

| | |
|---|---|
| I. | Bovine pancreatic trypsin inhibitor (Kunitz) family |
| II. | Pancreatic secretory trypsin inhibitor (Kazal) family |
| III. | *Streptomyces* subtilisin inhibitor family |
| IV. | Soybean trypsin inhibitor (Kunitz) family |
| V. | Soybean proteinase inhibitor (Bowman-Birk) family |
| VI. | Potato I inhibitor family |
| VII. | Potato II inhibitor family[a] |
| VIII. | *Ascaris* trypsin inhibitor family[a] |
| IX. | Other families |

[a] Information on families VII and VIII is marginal. It may be that establishment of these families is not yet justified.

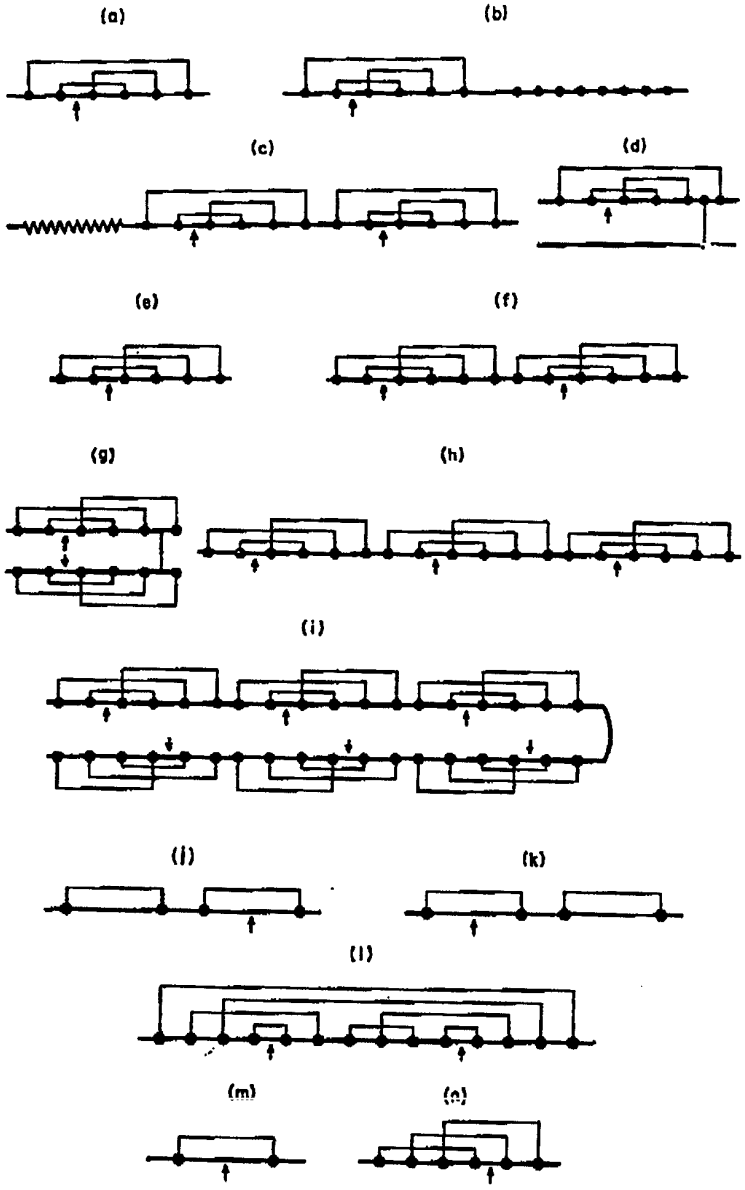


*Figure 1* Topological structures of proteinase inhibitors. Solid circle and arrow indicate, respectively, a half cystine residue and a reactive site. Inhibitors of the bovine pancreatic trypsin inhibitor (Kunitz) family obtained from: *a*, bovine pancreas, Kunitz, (148); *b*, red sea turtle egg white, chelonianin (149); *c*, human inter-$\alpha$-trypsin inhibitor (158, 159); *d*, B-chain of $\beta_1$ bungarotoxin (146).

Inhibitors of pancreatic secretory trypsin inhibitor (Kazal) family obtained from: *e*, bovine pancreas (138); *f*, canine submandibular glands (164, 165); *g*, red sea turtle egg white, *testudin* (Kato, I., Laskowski, M., Jr., unpublished results); *h*, avian egg white, ovomucoid (44); *i*, avian egg white ovoinhibitor (39).

Members of other inhibitor families obtained from: *j*, *Streptomyces albogriseolus* (137); *k*, soybean, Kunitz (167); *l*, lima bean, Bowman-Birk (168, 169); *m*, potato, I (177, 178); *n*, potato II (179–181).

and three-dimensional structure are known, it became a favorite for investigations by physical protein chemists, rivalling serum albumin and ribonuclease. It has been the subject of extensive studies on protein folding (188) (time course of formation of disulfide bridges), of extensive $^{1}H$ and $^{13}C$ nuclear magnetic resonance studies (189–191), and it is one of the favorite proteins for testing algorithms relating sequence to three-dimensional structure (192–198). It is worth pointing out that many other protein proteinase inhibitors share all the desirable properties of pancreatic trypsin inhibitor (Kunitz). In spite of extensive studies, the physiological function of this inhibitor is not known. It is present in virtually all bovine organs (38) (liver, lungs, parotid glands, spleen), but appears to be restricted only to bovids and caprids (199).

However, both cows and pigs have a homologous inhibitor, the colostrum trypsin inhibitor (200, 201). The bovine colostrum inhibitor was sequenced (150); it is a glycoprotein with a glycosylated Asn residue (202). Its physiological function is controversial; one suggestion is that it protects colostrum antibodies against tryptic digestion in the newborn (203). Evidence both for and against this suggestion has been adduced (204–206).

The inter-$\alpha$-trypsin inhibitor in mammalian serum is a huge molecule with a mol wt of 160,000. Treatment with an excess of trypsin liberates from it a COOH-terminal fragment of mol wt 14,000 (207). This fragment was sequenced from both bovine and human inter-$\alpha$-trypsin inhibitor (158, 159). It consists of two tandem domains, each homologous to bovine pancreatic trypsin inhibitor (Kunitz), connected by a short connecting peptide. The first of these domains (the penultimate domain of inter-$\alpha$-trypsin inhibitor) has Met (human) or Leu (bovine) as a $P_1$ residue. Surprisingly, it does not inhibit any of the enzymes tested against it. The second domain (the ultimate domain in inter-$\alpha$-trypsin inhibitor) has Arg as $P_1$ in both human and bovine material. It inhibits trypsin. The sequence preceding the two COOH terminal domains of inter-$\alpha$-trypsin inhibitor is not known; it may contain several more Kunitz type domains. Fragmentation of inter-$\alpha$-trypsin inhibitor also occurs in vivo. The fragments in serum are identified as acid-stable serum proteinase inhibitors. Because of their low molecular weights they are rapidly excreted and can be isolated from urine, when they are called mingin (208). Since many proteinase inhibitors consist of several independent domains connected by connecting peptides, these findings, so far unique, suggest that many smaller inhibitors will turn out to be fragments of larger ones. The presence of bovine pancreatic trypsin inhibitor (Kunitz), bovine colostrum inhibitor, and at least two Kunitz-type domains in inter-$\alpha$-trypsin inhibitor shows that the bovine genome codes for at least four Kunitz type inhibitor domains.

So far, no Kunitz type inhibitors have been reported from birds. However, snake venoms proved an unexpectedly rich source. Four different

single domain Kunitz type inhibitors (toxins I, K, B, and E) were isolated, and three of them (I, K, and E) were sequenced from the venom of the black mamba (154–156). The striking property of these is that toxin K ($P_1$ Lys) and toxin I ($P_1$ Tyr) do not inhibit trypsin and chymotrypsin, respectively, but toxin E ($P_1$ Lys) is a strong inhibitor of trypsin. Several Kunitz type inhibitors of trypsin were isolated and sequenced from venoms of several snakes (152, 153). $\beta_1$ Bungarotoxin (146) consists of two polypeptide chains held together by an interchain disulfide bridge. Chain A of 120 residues is homologous to phospholipase A, but chain B is homologous to Kunitz type inhibitors. However, chain B differs from the other members of the Kunitz group in two important ways. The residue corresponding to reactive site $P_1$ is deleted, thus suggesting that this is not an inhibitor, and there are seven (rather than the common six) half-cystine residues. The seventh residue is presumably responsible for the interchain disulfide. The $\beta_1$ bungarotoxin structure shows that Kunitz type inhibitor sequences may be covalently associated with protein sequences other than multiple Kunitz domains.

A much clearer example of this phenomenon is shown by the Kunitz type inhibitor from red sea turtle egg white, chelonianin (149). This inhibitor inhibits trypsin (presumably in its first domain where $P_1$ is Lys) and subtilisin (presumably on the second domain), independently. The first domain is closely homologous to Kunitz type inhibitors, while the second domain is not homologous to any of the inhibitor families described in this review and therefore presumably represents a new inhibitor family. Even though the subtilisin reactive site has not yet been assigned, it seems likely that the chelonianin is the first and, so far, the only example of a multidomain inhibitor, where the domains stem from different inhibitor families.

The Kunitz type inhibitors are not confined to vertebrates: several closely related inhibitors were isolated from the mucus of the garden snail (209, 210). Some of these have $P_1$ Arg, others $P_1$ Lys. One of them, K ($P_1$ Lys, 151), was sequenced. An inhibitor of the Kunitz class in sea anemone (211, 212) has also been partially sequenced (157). The gene coding for Kunitz type inhibitors is thus both very old and very widely distributed. It appears to have been present at the inception of the radiation of multicellular animals; in many animals it is present in several paralogous copies.

## PANCREATIC SECRETORY TRYPSIN INHIBITOR (KAZAL) FAMILY

The family is named after bovine pancreatic secretory trypsin inhibitor (Kazal) (213) which was found in all the vertebrates examined. The bovine (138), ovine (160), porcine (161), canine (K. Hochstrasser, personal communication), human (162), and turkey (163) inhibitors were sequenced. The inhibitors are stored in zymogen granules and secreted with the zymogens in the pancreatic juice. Consistent with their function is the clear specificity for inhibition of trypsin, failure to interact with trypsinogen, failure to

inhibit enterokinase (if they did, trypsinogen could not be activated in the duodenum) and other pancreatic serine proteinases of its own species (the inhibitor would then be diverted from its essential target, trypsin). These properties contrast with those of the pancreatic trypsin inhibitor (Kunitz). A large number of low-molecular-weight inhibitors were isolated from the sperm, seminal plasma, and seminal vesicles of various mammals (32). Most of these are referred to as acrosin inhibitors, since acrosin is a trypsin-like enzyme located in sperm acrosomes (215, 216). However, the physiological function of these inhibitors is far from clear. Two such inhibitors were sequenced; one from porcine seminal plasma completely and one from guinea pig seminal vesicles partially (166). Both are single domain inhibitors that are clearly homologous, but not identical, to Kazal pancreatic secretory inhibitors. It is highly likely that several other unsequenced inhibitors of this type also belong to the Kazal family on the basis of their amino acid analysis.

A double-headed inhibitor of trypsin (on one site) and of chymotrypsin, subtilisin, and elastase (on the other site) is present in extremely large amounts in dog submandibular glands (164, 214). This inhibitor consists of two tandem Kazal type domains, connected by a short connecting peptide (165). The first trypsin-inhibiting domain has $P_1$ Arg, the second $P_1$ Met. Dog seminal plasma inhibitor probably also exists and is another member of this class.

Ovomucoids, the major glycoprotein proteinase inhibitors from avian egg whites, have long been known, and chicken ovomucoid, a single-headed inhibitor of trypsin, has been widely studied (217). However, examination of inhibitory properties of ovomucoids from various birds produced a major surprise (218, 219). Some were single-headed (e.g. chicken), some double-headed (e.g. turkey, which has an independent site for trypsin and another for chymotrypsin, subtilisin, and elastase), and yet others were triple-headed (e.g. duck, with two sites for trypsin and one for chymotrypsin, subtilisin, and elastase). Furthermore, inhibitors from closely related birds have different specificities (e.g. golden pheasant ovomucoid inhibits chymotrypsin, while chicken ovomucoid inhibits trypsin). All of these findings were easier to understand when, upon sequencing, avian ovomucoids turned out to consist of three tandem, homologous (43, 44, 220), Kazal-type domains, in which the $P_1$ residue of the reactive site varies widely. The connecting peptide between the second and third domains can be readily hydrolyzed, and the resultant domain III and the double domain I-II are independently active. No evidence suggests any significant noncovalent interaction between the domains. The carbohydrate in ovomucoid is attached to four (in Japanese quail and turkey) or five (in chicken) Asn residues all in Asn-X-Thr/Ser sequences. The glycosylation of the site in the third domains is partial, therefore, after enzymatic hydrolysis of the

connecting peptide, two types of third domains are isolated by gel exclusion. Both have the same amino acid sequence but one is glycosylated, and the other is not. Both have identical inhibitory properties and behave identically in the low pH transition.

Along with other proteins synthesized by the oviduct, ovomucoids became objects of considerable interest to students of gene expression, protein synthesis, and secretion (221, 222). Chicken ovomucoid has a 22 residue signal peptide, which is removed from its $NH_2$ terminus during synthesis. Messenger RNA for chicken ovomucoid was isolated, and cDNA copy was prepared and cloned (223, 224). The sequencing of this material as well as of the DNA gene of ovomucoid is in progress.

Aside from ovomucoid, avian egg whites contain another inhibitor of serine proteinases, ovoinhibitor (41). This protein is present in much smaller and more variable amounts than ovomucoid (225). It is also a glycoprotein, but with a smaller carbohydrate content than ovomucoid (226). It has been extensively studied (227–232). Sequencing (almost complete) shows that Japanese quail ovoinhibitor consists of six tandem, homologous, Kazal type domains (39, 45). The $P_1$ residues of the six domains are, in order, Arg, Arg, Arg, Tyr, Met, and Met. It is striking that these three trypsin-type reactive sites are followed by three chymotrypsin-subtilisin-elastase-type reactive sites. A proteinase inhibitor with specificity identical to that of chicken ovoinhibitor and with almost identical molecular properties was isolated from chicken plasma (40). It was immunologically indistinguishable from ovoinhibitor (40), and had the same amino acid sequence for at least the first 30 residues from its amino terminal end (45). On these bases we believe that chicken ovoinhibitor and chicken plasma proteinase inhibitor are identical, the slight differences arising from differences in post-translational modification. It is thus likely that the principal function of ovoinhibitors is to serve as the major proteinase inhibitor in avian plasma (a physiological equivalent, but not a homologue of mammalian $\alpha_1$ proteinase inhibitors). On the basis of these studies birds code for at least ten paralogous Kazal type domains—six in ovoinhibitor, three in ovomucoid, and one in pancreatic secretory inhibitor. It is also likely that avian seminal plasma inhibitor contains a Kazal domain.

Turtle egg whites contain two proteinase inhibitors. Chelonianin, a hybrid of a Kunitz domain and another domain from a new family has already been described. The second is a two domain ovomucoid, testudin, which has two separate domains linked by one disulfide bridge.

Like Kunitz inhibitors, Kazal inhibitors are not limited to vertebrates. Crude extracts from leeches, *Hirudo medicinalis*, contain numerous trypsin inhibitors, called bdellins (233, 234). One of these, bdellin B-3, was partially sequenced. It is clearly a single-domain Kazal inhibitor (141).

STREPTOMYCES SUBTILISIN INHIBITOR FAMILY Various species of *Streptomyces* excrete into the medium protein proteinase inhibitors, with different specificities dependent upon species. One, called S-SI, from *Streptomyces albogriseolus,* inhibits subtilisin very strongly and chymotrypsin and trypsin very weakly (235, 236). It was prepared in high purity and large amount and was chosen as an object for collaborative study by several Japanese research teams. The amino acid sequence was obtained (137), the reactive site assigned, the three-dimensional structure was determined for the free inhibitor (237) and is in progress for the subtilisin-inhibitor complex (114, 238). A large number of physicochemical studies on the free inhibitor and on complex formation were made or are in progress (239–242). *Streptomyces antiplasminolyticus* excretes a potent plasmin inhibitor plasminostreptin. This inhibitor was also extensively characterized (136, 243, 244). The two inhibitors are strongly (70%) homologous. They consist of 113 and 109 residues respectively, each with two intrachain disulfide bridges. The reactive site is located in the second of these. S-SI forms strong noncovalent dimers and dimers inhibit two enzyme molecules without dissociation.

A striking observation is the pronounced sequence similarity in the reactive site region between S-SI and plasminostreptin on one hand and the Kazal family of inhibitors. In the region where it holds this similarity is sufficient to be called a homology; however, the remainder of the molecules appear unrelated. We are unable to decide whether the S-SI family and the Kazal family are products of divergent or of convergent evolution, and therefore, whether S-SI family should be a subgroup of the Kazal family or a separate family. In either case, the strong sequence similarity should make the comparison between them particularly useful in structure-to-function studies.

More *Streptomyces* inhibitors are being isolated and studied. They may prove to be as good or better than avian ovomucoid third domains as a demonstration of hypervariability of the reactive site between closely related species.

BOWMAN-BIRK INHIBITOR FAMILY These inhibitors are readily isolated from the seeds of all leguminous plants. A large number of them were briefly described (11), and a considerable number were sequenced (168–176). All of the sequenced ones consist of two tandem homology regions on the same polypeptide chain, each with a reactive site (245) (Table 3). However, a significant difference exists between the Bowman-Birk inhibitors, where homology regions are linked by interhomology region disulfide bonds (246), and the animal inhibitors from Kunitz pancreatic trypsin-inhibitor related family and Kazal secretory inhibitor family, where there are no interhomology region [inhibitor family, where there are no interhomology region (interdomain) disulfide bonds] (43). In spite of the pres-

ence of interhomology region disulfides, Ikenaka and co-workers split the two homology regions by making two specific peptide bond cleavages (CNBr and a pepsin-catalyzed cut) (247, 248). Both homology regions are separately active, but are weaker inhibitors than the parent molecule, presumably due to partial loss of rigidity. Synthetic decapeptides corresponding to the heterodetic cyclopeptide containing the reactive site were made (249). These are even more weakly inhibitory than the separated homology regions, but, nonetheless, they appear to serve as promising models for the study of the role of individual amino acid residues in inhibition.

In most Bowman-Birk type inhibitors the $P_1$ residue in the first ($NH_2$-terminal) homology region is Lys, and trypsin is inhibited, while the $P_1$ residue in the second homology region is Leu, and chymotrypsin is inhibited. However, these residues vary a good deal. Particularly interesting is the soybean inhibitor D-II, in which both $P_1$ residues are Arg and which is double headed for trypsin. This inhibitor can be viewed as an ancestor of other inhibitors (171). Soybean inhibitor C-II (172) and garden bean inhibitor II (173), both of which have Ala and Arg as their $P_1$ residues, inhibit both elastase and trypsin.

Many homologous Bowman-Birk inhibitors are isolated from each species of beans or even from a single bean (171). Some of these forms appear to differ only in the length of their $NH_2$-terminal sequence, but many represent significant differences in sequence. Their tendency to form homo- or hetero- dimers and trimers complicates the estimate of the number of forms present. However, Ikenaka et al isolated and sequenced three truly different inhibitors of this type from soybeans (171, 172).

SOYBEAN TRYPSIN INHIBITOR (KUNITZ) FAMILY The first plant inhibitor to be well characterized was soybean trypsin inhibitor (Kunitz). Its isolation and crystallization and that of its complex with trypsin by M. Kunitz is one of the classic achievements of inhibitor chemistry (185, 250). The sequence (167, 251, 252) and three-dimensional structure of the porcine trypsin-inhibitor complex were determined (112). This inhibitor served as the main substrate for work (largely in this laboratory) that established the standard mechanism and permitted development of semisynthetic reactive site replacement technique (104, 108). It is, therefore, surprising that this inhibitor stands rather alone—there are no reports of homologous inhibitors from the most common leguminous plants. However, a strikingly similar inhibitor from winged bean (253) was recently described, and inhibitors from rice (254) and from barley (255, 256) are probably homologous to soybean trypsin inhibitor (Kunitz).

Soybean trypsin inhibitor (Kunitz) is often confused with Bowman-Birk inhibitor, but the two are strikingly different. The polypeptide chain in single-headed Kunitz soybean inhibitor has 181 residues and only two

disulfide bridges (see Figure 1, *k*) while Bowman-Birk is typically double headed, has 7 disulfides, and only about 70 residues (Figure 1, *l*).

OTHER INHIBITOR FAMILIES We recently proposed (45) a list of inhibitor families in which we assigned the trypsin inhibitor(s) and chymotrypsin inhibitor(s) from *Ascaris* to two new separate families. It is now likely, but not certain, that one new family will suffice. *Ascaris lumbricoides* is a large parasitic worm, one type of which infests man and another pigs. These worms contain inhibitors for all of the digestive enzymes of the host (67). Pepsin and carboxypeptidase inhibitors from this source are listed in Table 1. One or more inhibitors of trypsin and several related inhibitors of chymotrypsin have been reported. The trypsin inhibitor isolated from *Ascaris* that infest pigs is an inefficient inhibitor of human trypsin, thus raising the possibility that this property may account for the inability of this species to infest humans. The sequence of this inhibitor has long been known (257), but the position of its reactive site was only recently assigned (R. J. Peanasky, W. J. Kohr, and M. Laskowski, Jr., unpublished results); in the process, the sequence was revised.

The chymotrypsin inhibitors of *Ascaris* appear to inhibit chymotrypsin, elastase, and subtilisin at the same reactive site (67). This behavior is similar to that of Kazal-type inhibitors (e.g. turkey ovomucoid third domain), but is in contrast to that of chymotryptic sites of Bowman-Birk type inhibitors, where only chymotrypsin is inhibited. Recently, both the preliminary sequence (258) of several *Ascaris* chymotrypsin inhibitors and the location of the reactive site (259) were announced.

Another extremely rich source of proteinase inhibitors are potatoes and related plants (9, 11). Many inhibitors, including carboxypeptidase inhibitor (Table 1) and several families of serine proteinase inhibitors, are present. The best understood are the potato I inhibitors, which are noncovalent tetramers of a single-chain inhibitor with only a single intrachain disulfide bridge (260, 261). Far more complex is the potato II family, where amino acid sequences have been reported for a "low-molecular-weight chymotrypsin" inhibitor (262), for an inhibitor from eggplant (263), and for inhibitory fragments of two larger serine proteinase inhibitors, IIa (264, 265) and IIb (180, 266). When all of these are aligned, homology is apparent. However, in contrast to other inhibitors, reactive sites (267, 268) do not align and there are many postulated deletions. At the present stage, therefore, the designation of a potato II family may be premature.

At least 100 small inhibitors not reported in this review have been described for serine proteinases and in some cases rather intensively studied. It seems premature to place them into families, but it is clear that new families will have to be created for some of them. The benefits in obtaining

enough information to allow for a family assignment will be quite large, since comparison of inhibitory specificities or of various thermodynamic and kinetic parameters of enzyme-inhibitor interaction is far more valid within a family.

INHIBITORS IN MAMMALIAN BLOOD PLASMA Many proteolytic processes (e.g. blood coagulation, blood clot dissolution, formation and destruction of peptide hormones) occur in blood plasma and require exquisite control. It is not surprising, therefore, that plasma contains many kinds of proteinase inhibitors—some in very high concentration. The principal ones in human plasma are listed below.

Of the inhibitors listed in Table 5, $\alpha_2$ macroglobulin is a general endopeptidase trap and its postulated mechanism of action has already been discussed. Inter-$\alpha$-trypsin inhibitor chain terminates in two domains belonging to the pancreatic trypsin inhibitor (Kunitz) family, the presumed reactive site lies in the COOH terminal domain. The other inhibitors listed in Table 5 do not appear to belong to any of the families discussed thus far. Furthermore, the sequence is not completely known for any of them and their mechanism of interaction with serine proteinase—in spite of a large body of literature dealing with this subject—is either unknown or highly controversial. This comparative lack of information results from the comparatively large size (required of all plasma proteins) and high lability of these inhibitors, which complicate isolation and experiments that are readily done with small stable inhibitors. As a result we resist detailed review and offer only thumbnail sketches without thorough analysis.

By far the most studied is $\alpha_1$-proteinase inhibitor (often called $\alpha_1$ antitrypsin—a bad name) because its genetic deficiency strongly predisposes individuals who are homozygous for its Z allele (in contrast to the normal M) to emphysema (270) and to liver disease. The Z inhibitor is fully active, but it is present in serum at 20% of the normal concentration. Partial sequencing shows that Z differs from M by the substitution of Lys for Glu in one of the internal sequence positions (271–273). In contrast to small inhibitors discussed earlier, $\alpha_1$-proteinase inhibitor contains no intrachain disulfide bridges (131). $\alpha_1$-Proteinase inhibitor inhibits an exceptionally broad spectrum of serine proteinases (e.g. trypsin, chymotrypsin, and elastase). A complex with one of these enzymes does not inhibit any of the others. This observation may indicate that all of the enzymes act at the same reactive site; a number of alternative explanations are also possible. If a single reactive site is accepted then good, but not compelling, evidence (182, 274) was presented that the site is a ... Met-Ser ... bond in the $NH_2$-terminal portion of the molecule (the sequence surrounding the reactive site is shown in Table 3).

Table 5 Principal proteinase inhibitors in human blood plasma (269)

| Name | Concentration mg/100 ml | Mol wt | Number of polypeptide chains | Heads[a] |
|---|---|---|---|---|
| $\alpha_1$-proteinase inhibitor | 290 ± 45 | 52,000 | 1 | 1 |
| $\alpha_1$ antichymotrypsin | 49 ± 7 | 69,000 | 1 | 1 |
| inter-$\alpha$-trypsin inhibitor | 50 | 160,000 | 1 | 2 |
| $\alpha_2$ antiplasmin | 7 ± 1 | 70,000 | 1 | 1 |
| antithrombin III | 24 ± 2 | 65,000 | 1 | 1 |
| $C_1$-inactivator | 24 ± 3 | 70,000 | 1 | 1 |
| $\alpha_2$-macroglobulin | 260 ± 70 | 720,000 | 4 | 1–2 or more |

[a] Maximal number of enzyme molecules simultaneously inhibited by one inhibitor molecule.

The most recently discovered (275–277) of the major plasma inhibitors is $\alpha_2$ antiplasmin. Antiplasmin reacts slowly with several trypsin-like enzymes but its reaction with plasmin is exceptionally fast. However, its rate of association is reduced (278) to more typical values either by enzymatic removal of that portion of the plasmin heavy (A) chain containing the strong lysine binding site or by addition of moderately low concentrations of lysine or of $\epsilon$-aminocaproic acid. The plasmin-$\alpha_2$ antiplasmin-inhibitor interaction is speeded by specific interaction between the inhibitor and the strong lysine binding site on the heavy chain of plasmin (278).

Antithrombin III is the major thrombin-directed inhibitor in plasma. The thrombin-antithrombin reaction is moderately slow in absence of added heparin; addition of heparin greatly accelerates the reaction. The current consensus appears to be that it is the antithrombin-heparin complex that reacts with thrombin very rapidly (279–281). Once the enzyme-inhibitor-heparin complex is formed, most of the heparin is released, and is therefore free to combine with additional antithrombin to further "catalyze" the enzyme-inhibitor reaction (282, 283). An essentially complete amino acid sequence of antithrombin III was recently reported (284), but the thrombin reactive site cannot be easily identified. Based on only a small portion of the $\alpha_1$ proteinase inhibitor sequence, it appears that antithrombin III and $\alpha_1$ proteinase inhibitor may be homologous.

## Acknowledgment

We are grateful to the National Institutes of Health for their long term support of protein-proteinase inhibitor research in our laboratory (NIH GM 10831), and to numerous colleagues around the world with whom we exchange new inhibitor information.

*Literature Cited*


1. Vogel, R., Trautschold, I., Werle, E. 1968. *Natural Proteinase Inhibitors.* New York:Academic. 159 pp.
2. Laskowski, M. Jr., Sealock, R. W. 1971. *The Enzymes* 3:375–473
3. Tschesche, H. 1974. *Angew. Chem. Int. Ed. Engl.* 13:10–28
4. Means, G. E., Ryan, D. S., Feeney, R. E. 1974. *Acc. Chem. Res.* 74:315–20
5. Kassell, B., Williams, M. J. 1976. In *Handbook of Biochemistry and Molecular Biology,* ed. G. D. Fasman, 2:583–668. Cleveland:CRC. 790 pp.
6. Perlmann, G. E., Lorand, L., eds. 1970. *Methods Enzymol.* 19:1–1042
7. Lorand, L., ed. 1976. *Methods Enzymol.* 45(Pt. B.):1–939
8. Mihalyi, E. 1978. *Application of Proteolytic Enzymes to Protein Structure Studies,* Vol. 1. West Palm Beach:CRC. 328 pp.
9. Ryan, C. A. 1973. *Ann. Rev. Plant Physiol.* 24:173–96
10. Odani, S., Ikenaka, T. 1977. *Seikagaku* 49:1–20 (In Japanese)
11. Richardson, M. 1977. *Phytochemistry* 16:159–69
12. Mitsui, Y., Watanabe, Y., Hirono, S. 1979. *Tanpakushitsu Kakusan Koso* 24:96–118 (In Japanese)
13. Kato, I. 1979. *Tanpakushitsu Kakusan Koso* 24:667–79 (In Japanese)
14. Fritz, H., Tschesche, H., eds. 1971. *Proc. Ist Int. Res. Conf. Proteinase Inhibitors,* Berlin:Walter de Gruyter. 304 pp.
15. Fritz, H., Tschesche, H., Greene, L. J., Truscheit, E., eds. 1974. *Bayer Symp. V, Proteinase Inhibitors,* Berlin:Springer. 751 pp.
16. Peeters, H., ed. 1976. *Protides of the Biological Fluids, 23rd,* Oxford: Pergamon. 699 pp.
17. Reich, E., Rifkin, D. B., Shaw, E., eds. 1975. *Cold Spring Harbor Conf. Cell Proliferation,* Vol. 2, *Proteases and Biological Control,* Cold Spring Harbor Lab. 1021. pp.
18. Magnusson, S., Ottesen, M., Foltmann, B., Danø, K., Neurath, H., eds. 1978. *Proc. 11th FEBS Meet. Regulatory Proteolytic Enzymes and Their Inhibitors.* Oxford: Pergamon. 236 pp.
19. Li, C. H., ed. 1978. *Proc. Int. Symp. Proteins: Versatility of Proteins,* New York:Academic. 465 pp.
20. Barrett, A. J., ed. 1977. *Proteinases in Mammalian Cells and Tissues,* Amsterdam: Elsevier/North-Holland. 735 pp.
21. Collen, D., Wiman, B., Verstraete, M., eds. 1979. *The Physiological Inhibitors of Blood Coagulation and Fibrinolysis,* Amsterdam:Elsevier/North-Holland. 295 pp.
22. Cohen, G. N., Holzer, H., eds. 1979. *Limited Proteolysis in Microorganisms,* Wash. DC: USGPO 255 pp.
23. Shainkin, R., Birk, Y. 1970. *Biochim. Biophys. Acta* 221:502–13
24. Lindberg, M. U., Skoog, L. 1970. *Eur. J. Biochem.* 13:326–35
25. Baraganca, B. M., Sambray, Y. M., Sambray, R. Y. 1970. *Eur. J. Biochem.* 13:410–15
26. Walsh, D. A., Ashby, C. D., Gonzalez, C., Calkins, D., Fischer, E. H., Krebs, E. G. 1971. *J. Biol. Chem.* 246:1977–85
27. Jeppsson, J.-O., Laurell, C.-B. 1975. See Ref. 17, pp. 405–14
28. Davie, E. W., Fujikawa, K., Legaz, M. E., Kato, H. 1975. See Ref. 17, pp. 65–77
29. Neurath, H., Walsh, K. A., 1978. See Ref. 18, pp. 1–14
30. Green, T. R., Ryan, C. A. 1972. *Science* 175:776–77
31. Ryan, C. A. 1974. See Ref. 15, pp. 667–73
32. Fritz, H., Schiessler, H., Schill, W.-B., Tschesche, H., Heimburger, N., Wallner, O. 1975. See Ref. 17, pp. 737–66
33. Lenney, J. F., Tolan, J. R., Lee, A. G. 1979. See Ref. 22, pp. 21–25
34. Holzer, H. 1979. See Ref. 22. pp. 59–60
35. Murao, S., Sato, S. 1972. *Agr. Biol. Chem.* 36:160–63
36. Kakinuma, A., Sugino, H., Moriya, N., Isono, M. 1978. *J. Biol. Chem.* 253:1529–37
37. Ikenaka, T., Odani, S. 1978. See Ref. 18, pp. 207–16
38. Fritz, H., Kruck, I., Russe, I., Liebich, H. G. 1979. *Hoppe-Seylers Z. Physiol. Chem.* 360:437–44
39. Kato, I., Kohr, W. J. 1978. *Fed. Proc.* 37:1337 (Abstr.)
40. Barrett, A. J. 1974. *Biochim. Biophys. Acta* 371:52–62
41. Matsushima, K. 1958. *Science* 127:1178–79
42. Johnson, D. A., Travis, J. 1976. *Biochem. Biophys. Res. Commun.* 72: 33–39
43. Kato, I., Schrode, J., Wilson, K. A., Laskowski, M. Jr. 1976. See Ref. 16, pp. 235–43
44. Kato, I., Kohr, W. J., Laskowski, M. Jr. 1978. See Ref. 18, pp. 197–206
45. Laskowski, M. Jr., Kato, I., Kohr, W. J. 1978. See Ref. 19, pp. 307–18
46. Harpel, P. C., Mosesson, M. W.,

Cooper, N. R. 1975. See Ref. 17, pp. 387–404
47. Hall, P. K., Roberts, R. C. 1978. *Biochem. J.* 173:27–38
48. Haverback, B. J., Dyce, B., Baundy, H. F. 1962. *J. Clin. Invest.* 41:972–80
49. Eddeland, A., Ohlsson, K. 1978. *Hoppe-Seylers Z. Physiol. Chem.* 359:379–84
50. Hartley, B. S. 1960. *Ann. Rev. Biochem.* 29:45–72
51. Barrett, A. J., Starkey, P. M. 1973. *Biochem. J.* 133:709–24
52. Balldin, G., Laurell, C.-B., Ohlsson, K. 1978. *Hoppe-Seylers Z. Physiol. Chem.* 359:699–708
53. Fossum, K., Whitaker, J. R. 1968. *Arch. Biochem. Biophys.* 125:367–75
54. Sen, L. C., Whitaker, J. R. 1973. *Arch. Biochem. Biophys.* 158:623–32
55. Keilova, H., Tomasek, V. 1975. *Collection Czechoslov. Chem. Commun.* 40:218–24
56. Keilova, H., Tomasek, V. 1974. *Biochim. Biophys. Acta* 334:179–86
57. Reddy, M. N., Keim, P. S., Heinrikson, R. L., Kezdy, F. J. 1975. *J. Biol. Chem.* 250:1741–50
58. Jarvinen, M. 1976. *Acta Chem. Scand. B* 30:933–40
59. Udaka, K., Hayashi, H. 1965. *Biochim. Biophys. Acta* 97:251–61
60. Udaka, K., Hayashi, H. 1965. *Biochim. Biophys. Acta* 104:600–3
61. Goldstein, Z., Trop, M., Birk, Y. 1973. *Nature New Biol.* 246:29–31
62. Kopitar, M., Brzin, J., Zvonar, T., Locnikar, P., Kregar, I., Turk, V. 1978. *FEBS Lett.* 91:355–59
63. Roughley, P. J., Murphy, G., Barrett, A. J. 1978. *Biochem. J.* 169:721–24
64. Sasaki, M., Minakata, K., Yamamoto, H., Niwa, M., Kato, T., Ito, N. 1977. *Biochem. Biophys. Res. Commun.* 76:917–24
65. Ryley, H. C. 1979. *Biochem. Biophys. Res. Commun.* 89:871–78
66. Starkey, P. M., Barrett, A. J. 1973. *Biochem. J.* 131:823–31
67. Peanasky, R. J., Abu-Erreish, G. M., Gaush, C. R., Homandberg, G. A., O'-Heeron, D., Linkenheil, R. K., Kucich, U., Babin, D. R. 1974. See Ref. 15, pp. 649–66
68. Abu-Erreish, G. M., Peanasky, R. J. 1974. *J. Biol. Chem.* 249:1558–65
69. Abu-Erreish, G. M., Peanasky, R. J. 1974. *J. Biol. Chem.* 249:1566–71
70. Keilova, H., Tomasek, V. 1972. *Biochim. Biophys. Acta* 284:461–64
71. Keilova, H., Tomasek, V. 1976. *Collection Czechoslov. Chem. Commun.* 41:489–97
72. Keilova, H., Tomasek, V. 1976. *Collection Czechoslov. Chem. Commun.* 41:2440–47
73. Sakato, K., Tanaka, H., Misawa, M. 1975. *Eur. J. Biochem.* 55:211–19
74. Kuettner, K. E., Hiti, J., Eisenstein, R. Harper, E. 1976. *Biochem. Biophys. Res. Commun.* 72:40–46
75. Kuettner, K. E., Soble, L., Croxen, R. L., Marczynska, B., Hiti, J., Harper, E. 1977. *Science* 196:653–54
76. Murphy, G., Cartwright, E. C., Sellers, A., Reynolds, J. J. 1977. *Biochim. Biophys. Acta* 483:493–98
77. Nolan, J. C., Ridge, S., Oronsky, A. L., Slakey, L. L., Kerwar, S. S. 1978. *Biochem. Biophys. Res. Commun.* 83:1183–90
78. Woolley, D. E., Roberts, D. R., Evanson, J. M. 1976. *Nature* 261:325–27
79. Nishiura, I., Tanaka, K., Yamato, S., Murachi, T. 1978. *J. Biochem. Tokyo* 84:1657–59
80. Sellers, A., Murphy, G., Meikle, M. C., Reynolds, J. J. 1979. *Biochem. Biophys. Res Commun.* 84:581–87
81. Ryan, C. A., Hass, G. M., Kuhn, R. W., Neurath, H. 1974. *J. Biol. Chem.* 249:5495–99
82. Hass, G. M., Nau, H., Biemann, K., Grahn, D. T., Ericsson, L. H., Neurath, H. 1975. *Biochemistry* 14:1334–42
83. Hass, G. M., Ako, H., Grahn, D. T., Neurath, H. 1976. *Biochemistry* 15:93–100
84. Ako, H., Hass, G. M., Grahn, D. T., Neurath, H. 1976. *Biochemistry* 15:2573–78
85. Leary, T. R., Grahn, D. T., Neurath, H., Hass, G. M. 1979. *Biochemistry* 18:2252–56
86. Homandberg, G. A., Peanasky, R. J. 1976. See Ref. 16, pp. 279–84
87. Kaehn, K., Kula, M.-R. 1977. *FEBS Lett.* 79:85–90
88. Yu, P. H., Kula, M. R., Tsai, H. 1976. *Can. J. Biochem.* 54:699–703
89. Bagdy, D., Barabas, E., Graf, L., Petersen, T. E., Magnusson, S. 1976. See Ref. 7, pp. 669–78
90. Ozawa, K., Laskowski, M. Jr. 1966. *J. Biol. Chem.* 241:3955–61
91. Finkenstadt, W. R., Hamid, M. A., Mattis, J. A., Schrode, J., Sealock, R. W., Wang, D., Laskowski, M. Jr. 1974. See Ref. 15, pp. 389–411
92. Estell, D. A., Wilson, K. A., Laskowski, M. Jr. 1980. *Biochemistry* 19:131–37
93. Niekamp, C. W., Hixson, H. F. Jr., Las-

kowski, M. Jr. 1969. *Biochemistry* 8:16–22
94. Mattis, J. A., Laskowski, M. Jr. 1973. *Biochemistry* 12:2239–44
95. Krahn, J., Stevens, F. C. 1973. *Biochemistry* 12:1330–35
96. Sealock, R. W., Laskowski, M. Jr. 1973. *Biochemistry* 12:3139–46
97. Schrode, J. 1975. *The three domains in the chicken ovomucoid molecule and the location and properties of its reactive site against trypsin.* PhD thesis. Purdue Univ.
98. Tschesche, H., Kupfer, S. 1976. *Z. Physiol. Chem.* 357:769–76
99. Finkenstadt, W. R., Laskowski, M. Jr. 1965. *J. Biol. Chem.* 240:PC962–63
100. Kowalski, D., Laskowski, M. Jr. 1972. *Biochemistry* 11:3451–59
101. Finkenstadt, W. R., Laskowski, M. Jr. 1967. *J. Biol. Chem.* 242:771–73
102. Hixson, H. F. Jr., Laskowski, M. Jr. 1970. *J. Biol. Chem.* 245:2027–35
103. Laskowski, M. Jr., Finkenstadt, W. R., Mattis, J. A., McKee, R. E. 1976. See Ref. 16, pp. 85–91
104. Sealock, R. W., Laskowski, M. Jr. 1969. *Biochemistry* 8:3703–10
105. Leary, T. R., Laskowski, M. Jr. 1973. *Fed. Proc.* 32:465 (Abstr.)
106. Kowalski, D., Leary, T. R., McKee, R. E., Sealock, R. W., Wang, D., Laskowski, M. Jr. 1974. See Ref. 15, pp. 311–24
107. Jering, H., Tschesche, H. 1976. *Eur. J. Biochem.* 61:453–63
108. Kowalski, D., Laskowski, M. Jr. 1976. *Biochemistry* 15:1300–9
109. Odani, S., Ikenaka, T. 1978. *J. Biochem. Tokyo* 84:1–9
110. Luthy, J. A., Praissman, M., Finkenstadt, W. R., Laskowski, M. Jr. 1973. *J. Biol. Chem.* 248:1760–71
111. Quast, U., Engel, J., Steffen, E., Tschesche, H., Kupfer, S. 1978. *Eur. J. Biochem.* 86:353–60
112. Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H., Blow, D. M. 1974. *Biochemistry* 13:4212–28
113. Satow, Y., Mitsui, Y., Iitaka, Y. 1978. *J. Biochem. Tokyo* 84:897–906
114. Mitsui, Y., Satow, Y., Wanatabe, Y., Hirono, S., Iitaka, Y. 1979. *Nature* 277:447–52
115. Huber, R., Kukla, D., Ruhlmann, A., Epp, O., Formanek, H. 1970. *Naturwissenschaften* 57:389–92
116. Deisenhofer, J., Steigemann, W. 1975. *Acta Cryst. B* 31:238–50
117. Ruhlmann, A., Kukla, D., Schwager, P., Bartels, K., Huber, R. 1973. *J. Mol. Biol.* 77:417–36
118. Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J., Steigemann, W. 1974. *J. Mol. Biol.* 89:73–101
119. Huber, R., Bode, W., Kukla, D., Kohl, U., Ryan, C. A. 1975. *Biophys. Struct. Mechanism* 1:189–201
120. Bode, W., Schwager, P., Huber, R. 1978. *J. Mol. Bio.* 118:99–112
121. Huber, R., Bode, W. 1978. See Ref. 18, pp. 15–34
122. Huber, R., Bode, W. 1978. *Acc. Chem. Res.* 11:114–122
123. Bode, W. 1979. See Ref. 21, pp. 5–16
124. Ako, H., Foster, R. J., Ryan, C. A. 1974. *Biochemistry* 13:132–39
125. Kraut, J. 1977. *Ann. Rev. Biochem.* 46:331–58
126. Neves, D. E., Markley, J. L., Welch, M. E., Laskowski, M. Jr. 1979. *Fed. Proc.* 38:474 (Abstr.)
127. Hunkapiller, M. W., Forgac, M. D., Yu, E. H., Richards, J. H. 1979. *Biochem. Biophys. Res. Commun.* 87: 25–31
128. Robertus, J. D., Kraut, J., Alden, R. A., Birktoft, J. J. 1972. *Biochemistry* 11: 4293–304
129. Laskowski, M. Jr., Kato, I., Leary, T. R., Schrode, J., Sealock, R. W. 1974. See Ref. 15, pp. 597–611
130. Seemuller, U., Meier, M., Ohlsson, K., Muller, H.-P., Fritz, H. 1977. *Hoppe-Seylers Z. Physiol. Chem.* 358:1105–17
131. Crawford, I. P. 1973. *Arch. Biochem. Biophys.* 156:215–22
132. Wilson, K. A., Laskowski, M. Sr. 1974. See Ref. 15, pp. 286–90
133. Rigbi, M. 1971. See Ref. 14, pp. 74–88
134. Biddlingmeyer, U. D. V., Leary, T. R., Laskowski, M. Jr. 1972. *Biochemistry* 11:3303–10
135. Estell, D. A., Laskowski, M. Jr. 1980. *Biochemistry* 19:124–31
136. Sugino, H., Kakinuma, A., Iwanaga, S. 1978. *J. Biol. Chem.* 253:1546–55
137. Ikenaka, T., Odani, S., Sakai, M., Nabeshima, Y., Sato, S., Murao, S. 1974. *J. Biochem. Tokyo* 76:1191–1209
138. Greene, L. J., Bartelt, D. C. 1969. *J. Biol. Chem.* 244:2646–57
139. Bartelt, D. C., Greene, L. J. 1971. *J. Biol. Chem.* 246:2218–29
140. Tan, N. H., Kaiser, E. T. 1977. *Biochemistry* 16:1531–41
141. Krejci, K., Fritz, H. 1976. *FEBS Lett.* 64:152–55
142. Kowalski, D., Laskowski, M. Jr. 1976. *Biochemistry* 15:1309–15
143. Rhodes, M. B., Bennett, N., Feeney, R. E. 1960. *J. Biol. Chem.* 235:1686–93

144. Sato, S., Murao, S. 1974. *Agr. Biol. Chem.* 36:1734–44
145. Ryan, C. A., Shumway, L. K. 1971. See Ref. 14, pp. 175–88
146. Kondo, K., Narita, K., Lee, C.-Y. 1978. *J. Biochem. Tokyo* 83:101–15
147. Stevens, F. C. 1971. See Ref. 14, pp. 149–55
148. Kassell, B., Radicevic, M., Ansfield, M. J., Laskowski, M. Sr. 1965. *Biochem. Biophys. Res. Commun.* 18:255–58
149. Kato, I., Tominaga, N. 1979. *Fed. Proc.* 38(3):832 (Abstr.)
150. Cechova, D., Jonakova, V., Sorm, F. 1971. *Collect. Czech. Chem. Commun.* 36:3342–57
151. Tschesche, H., Dietl, T. 1975. *Eur. J. Biochem.* 58:439–51
152. Takahashi, H., Iwanaga, S., Kitagawa, T., Hokama, Y., Suzuki, T. 1974. *J. Biochem. Tokyo* 76:721–33
153. Hokama, Y., Iwanaga, S., Tatsuki, T., Suzuki, T. 1976. *J. Biochem. Tokyo* 79:559–78
154. Joubert, F. J., Strydom, D. J. 1978. *Eur. J. Biochem.* 87:191–98
155. Strydom, D. J. 1977. *Biochim. Biophys. Acta* 494:361–69
156. Strydom, D. J. 1973. *Nature New Biol.* 243:88–89
157. Wunderer, G., Beress, L., Machleidt, W., Fritz, H. 1976. See Ref. 16, pp. 285–88
158. Hochstrasser, K., Wachter, E., Bretzel, G. 1976. *Hoppe-Seylers Z. Physiol. Chem.* 357:1659–61
159. Hochstrasser, K., Wachter, E., Bretzel, G. 1978. See Ref. 18, pp. 225–34
160. Hochstrasser, K., Schramm, W., Fritz, H., Schwartz, S., Werle, E. 1969. *Hoppe-Seylers Z. Physiol. Chem.* 350:893–96
161. Tschesche, H., Wachter, E. 1970. *Eur. J. Biochem.* 16:187–98
162. Bartelt, D. C., Shapanka, R., Green, L. J. 1977. *Arch. Biochem. Biophys.* 179:189–99
163. Bogard, W. C. Jr., Laskowski, M. Jr. 1979. *Fed Proc.* 38(3):836 (Abstr.)
164. Hochstrasser, K., Bretzel, G., Wachter, E. 1976. See Ref. 16, pp. 267–70
165. Hochstrasser, K., Fritz, H. 1975. *Hoppe-Seylers Z. Physiol. Chem.* 356:1659–62
166. Tschesche, H., Kupfer, S., Klauser, R., Fink, E., Fritz, H. 1976. See Ref. 16, pp. 255–66
167. Koide, T., Ikenaka, T. 1973. *Eur. J. Biochem.* 32:417–31
168. Tan, C. G. L., Stevens, F. C. 1971. *Eur. J. Biochem.* 18:515–23
169. Stevens, F. C., Wuerz, S., Krahn, J. 1974. See Ref. 15, pp. 344–54
170. Odani, S., Koide, T., Ikenaka, T. 1971. *Proc. J. Acad.* 47:621–24
171. Odani, S., Ikenaka, T. 1978. *J. Biochem. Tokyo* 83:737–45
172. Odani, S., Ikenaka, T. 1977. *J. Biochem. Tokyo* 82:1523–31
173. Wilson, K. A., Laskowski, M. Sr. 1975. *J. Biol. Chem.* 250:4261–67
174. Belew, M., Eaker, D. 1976. *Eur. J. Biochem.* 62:499–508
175. Ishikawa, C., Nakamura, S., Watanabe, K., Takahashi, K. 1979. *FEBS Lett.* 99:97–100
176. Abe, O., Shimokawa, Y., Araki, T., Kuromizu, K. 1978. *J. Biochem. Tokyo* 83:1749–56
177. Richardson, M. 1974. *Biochem. J.* 137:101–12
178. Richardson, M., Barker, R. D. J., McMillan, R. T., Cossins, L. M. 1977. *Phytochemistry* 16:837–39
179. Kiyohara, T., Fujii, M., Iwasaki, T., Yoshikawa, M. 1973. *J. Biochem. Tokyo* 74:675–82
180. Iwasaki, T., Kiyohara, T., Yoshikawa, M. 1976. *J. Biochem. Tokyo* 79:381–91
181. Iwasaki, T., Wada, J., Kiyohara, T., Yoshikawa, M. 1977. *J. Biochem. Tokyo* 82:991–1004
182. Johnson, D., Travis, J. 1978. *J. Biol. Chem.* 253:7142–44
183. Kress, L. F., Kurecki, T., Chan, S. K., Laskowski, M. Sr. 1979. *J. Biol. Chem.* 254:5317–20
184. Kunitz, M., Northrop, J. H. 1936. *J. Gen. Physiol.* 19:991–1007
185. Kunitz, M. 1947. *J. Gen. Physiol.* 30:311–20
186. Yajima, H., Kiso, Y., Okada, Y., Watanabe, H. 1974. *J. Chem. Soc. Chem. Comm.* 1974(3):106–7
187. Noda, K., Terada, S., Mitsuyasu, N., Waki, M., Kato, T., Izumiya, N. 1971. *Naturwissenschaften* 58:147–48
188. Creighton, T. E. 1978. *Prog. Molec. Biol.* 33:231–97
189. Wuthrich, K., Wagner, G. 1978. *Trends Biochem. Sci.* 3:227–30
190. Wagner, G., Wuthrich, K. 1978. *Nature* 275:247–48
191. Snyder, G. H., Rowan, R., Karplus, S., Sykes, B. D. 1975. *Biochemistry* 14:3765–77
192. Tanaka, S., Scheraga, H. A. 1975. *Proc. Natl. Acad. Sci. USA* 72:3802–6
193. Tanaka, S., Scheraga, H. A. 1977. *Proc. Natl. Acad. Sci. USA* 74:1320–23
194. Levitt, M., Warshel, A. 1975. *Nature* 253:694–98

195. Kuntz, I. D., Crippen, G. M., Kollman, P. A., Kimelman, D. 1976. *J. Mol. Biol.* 106:983–94
196. Robson, B., Osguthorpe, D. J. 1979. *J. Mol. Biol.* 132:19–51
197. McCammon, J. A., Wolynes, P. G., Karplus, M. 1979. *Biochemistry* 18:927–42
198. Chou, P. Y., Fasman, G. D. 1974. *Biochemistry* 13:222–45
199. Werle, E. 1964. *Z. Physiol. Chem.* 338:228–30
200. Laskowski, M. Jr., Mars, P. H., Laskowski, M. Sr. 1952. *J. Biol. Chem.* 198:745–52
201. Cechova, D., Jonakova-Svestkova, V., Sorm, F. 1970. *Collect. Czech. Chem. Commun.* 35:3085–91
202. Klauser, R., Cechova, D., Tschesche, H. 1978. *Hoppe-Seylers Z. Physiol. Chem.* 359:173–80
203. Laskowski, M. Jr., Laskowski, M. Sr. 1951. *J. Biol. Chem.* 190:563–68
204. Baintner, K. 1973. *Acta Vet. Acad. Sci. Hung.* 23:247–51
205. Baintner, K. 1973. *Acta Vet. Acad. Sci. Hung.* 23:279–84
206. Veselsky, L., Cechova, D., Jonakova, V. 1978. *Hoppe-Seylers Z. Physiol. Chem.* 359:873–78
207. Hochstrasser, K., Bretzel, G., Feuth, H., Hilla, W., Lempart, K. 1976. *Hoppe-Seylers Z. Physiol. Chem.* 357:153–62
208. Astrup, T., Nissen, U. 1964. *Nature* 203:255–57
209. Tschesche, H., Dietl, T. 1972. *Hoppe-Seylers Z. Physiol. Chem.* 353:1189–93
210. Dietl, T., Tschesche, H. 1974. See Ref. 15, pp. 254–64
211. Fritz, H., Brey, B., Beress, L. 1972. *Hoppe-Seylers Z. Physiol. Chem.* 353:19–30
212. Wunderer, G., Kummer, K., Fritz, H., Beress, L., Machleidt, W. 1974. See Ref. 15, pp. 277–81
213. Kazal, L. A., Spicer, D. S., Brahinsky, R. A. 1948. *J. Am. Chem. Soc.* 70:3034–40
214. Fritz, H., Jaumann, E., Meister, R., Pasquay, P., Hochstrasser, K., Fink, E. 1971. See Ref. 14, pp. 257–70
215. Zaneveld, L. J. D., Polakoski, K. L., Robertson, R. T., Williams, W. L. 1971. See Ref. 14, pp. 236–44
216. Zaneveld, L. J. D., Polakoski, K. L., Schumacher, G. F. B. 1975. See Ref. 17, pp. 683–706
217. Feeney, R. E., Allison, R. G. 1969. *Evolutionary Biochemistry of Proteins.* New York: Wiley-Interscience, 290 pp.
218. Rhodes, M. B., Bennett, N., Feeney, R. E. 1960. *J. Biol. Chem.* 235:1686–93
219. Osuga, D. T., Bigler, J. C., Uy, R. L., Sjøberg, L., Feeney, R. E. 1974. *Comp. Biochem. Physiol.* B48:519–33
220. Beeley, J. G. 1976. *Biochem. J.* 155: 345–51
221. Palmiter, R. D., Thibodeau, S. N., Gagnon, J., Walsh, K. A. 1978. See Ref. 18, pp. 89–101
222. Thibodeau, S. N., Palmiter, R. D., Walsh, K. A. 1978. *J. Biol. Chem.* 253:9018–23
223. Catterall, J. F., Stein, J. P., Lai, E. C., Woo, S. L. C., Dugaiczyk, A., Mace, M. L., Means, A. R., O'Malley, B. W. 1979. *Nature* 278:323–27
224. Nordstrom, J. L., Roop, D. R., Tsai, M.-J., O'Malley, B. W. 1979. *Nature* 278:328–31
225. Liu, W.-H., Means, G. E., Feeney, R. E. 1971. *Biochim. Biophys. Acta* 229: 176–85
226. Tomimatsu, Y., Clary, J. J., Bartulovich, J. J. 1966. *Arch. Biochem. Biophys.* 115:536–44
227. Davis, J. G., Zahnley, J. C., Donovan, J. W. 1969. *Biochemistry* 8:2044–53
228. Zahnley, J. C., Davis, J. G. 1970. *Biochemistry* 9:1428–33
229. Zahnley, J. C. 1974. *J. Biol. Chem.* 249:4282–85
230. Gertler, A., Feinstein, G. 1971. *Eur. J. Biochem.* 20:547–52
231. Feinstein, G., Gertler, A. 1972. *Eur. J. Biochem.* 31:25–31
232. Shechter, Y., Burstein, Y., Gertler, A. 1977. *Biochemistry* 16:992–97
233. Fritz, H., Oppitz, K. H., Gebhardt, M., Oppitz, I., Werle, E., Marx, R. 1969. *Hoppe-Seylers Z. Physiol. Chem.* 350:91–92
234. Fritz, H., Gebhardt, M., Meister, R., Fink, E. 1971. See Ref. 14, pp. 271–80
235. Murao, S., Sato, S., Muto, N. 1972. *Agr. Biol. Chem.* 36:1737–44
236. Sato, S., Murao, S. 1973. *Agr. Biol. Chem.* 37:1067–74
237. Mitsui, Y., Satow, Y., Sakamaki, T., Iitaka, Y. 1977. *J. Biochem.* 82:295–98
238. Mitsui, Y., Satow, Y., Watanabe, Y., Iitaka, Y. 1979. *J. Mol. Biol.* 131:697–724
239. Inouye, K., Tonomura, B., Hiromi, K. 1975. *Agr. Biol. Chem.* 39:1159–61
240. Inouye, K., Tonomura, B., Hiromi, K. 1979. *J. Biochem. Tokyo* 85:1115–26
241. Inouye, K., Tonomura, B., Hiromi, K., Fujiwara, K., Tsuru, D. 1979. *J. Biochem. Tokyo* 85:127–34
242. Inouye, K., Tonomura, B., Hiromi, K.

1979. *Arch. Biochem. Biophys.* 192: 260–69
243. Kakinuma, A., Sugino, H., Moriya, N., Isono, M. 1978. *J. Biol. Chem.* 253:1529–37
244. Sugino, H., Nakagawa, S., Kakinuma, A. 1978. *J. Biol. Chem.* 253:1538–45
245. Krahn, J., Stevens, F. C. 1970. *Biochemistry* 9:2646–52
246. Odani, S., Ikenaka, T. 1973. *J. Biochem. Tokyo* 74:697–715
247. Odani, S., Ikenaka, T. 1973. *J. Biochem. Tokyo* 74:857–60
248. Smirnoff, P., Khalef, S., Birk, Y., Applebaum, S. W. 1976. *Biochem. J.* 157:745–51
249. Terada, S., Sato, K., Kato, T., Izumiya, N. 1978. *FEBS Lett.* 90:89–92
250. Kunitz, M. 1947. *J. Gen. Physiol.* 30:291–310
251. Koide, T., Ikenaka, T. 1973. *Eur. J. Biochem.* 32:401–7
252. Koide, T., Tsunasawa, S., Ikenaka, T. 1973. *Eur. J. Biochem.* 32:408–16
253. Kortt, A. A. 1979. *Biochim. Biophys. Acta* 577:371–82
254. Kato, I., Tominaga, N., Kihara, H. 1972. *Symp. Protein Structures, 23rd, Maebashi,* pp. 53–56. (In Japanese)
255. Yoshikawa, M., Iwasaki, T., Fujii, M., Oogaki, M. 1976. *J. Biochem.* 79:765–73
256. Mikola, J., Suolinna, E.-M. 1971. *Arch. Biochem. Biophys.* 144:566–75
257. Fraefel, W., Archer, R. 1968. *Biochim. Biophys. Acta* 154:615–17
258. Rabin, D. 1979. *Fed. Proc.* 38:833 (Abstr.)
259. Peanasky, R. J., Homandberg, G. A., Minor, S., Bentz, Y. 1979. *Fed. Proc.* 38:833 (Abstr.)
260. Ryan, C. A. 1966. *Biochemistry* 5:1592–96
261. Melville, J. C., Ryan, C. A. 1972. *J. Biol. Chem.* 247:3445–53
262. Hass, G. M., Eddy, J., Gentry, L., Derr, J., Ryan, C. A. 1978. *Fed. Proc.* 37:1337 (Abstr.)
263. Richardson, M. 1979. *FEBS Lett.* 104:322–26
264. Iwasaki, T., Kiyohara, T., Yoshikawa, M. 1976. *J. Biochem. Tokyo* 79:381–91
265. Iwasaki, T., Kiyohara, T., Yoshikawa, M. 1974. *J. Biochem. Tokyo* 75:843–51
266. Iwasaki, T., Wada, J., Kiyohara, T., Yoshikawa, M. 1975. *J. Biochem. Tokyo* 78:1267–74
267. Iwasaki, T., Kiyohara, T., Yoshikawa, M. 1973. *J. Biochem. Tokyo* 73: 1039–48
268. Iwasaki, T., Kiyohara, T., Yoshikawa, M. 1973. *J. Biochem. Tokyo* 74:335–40
269. Collen, D., Wiman, B. 1979. See Ref. 21, pp. 3–4
270. Laurell, C.-B., Eriksson, S. 1963. *Scand. J. Clin Lab. Invest.* 15:132–40
271. Jeppsson, J.-O. 1976. *FEBS Lett.* 65:195–97
272. Owen, M. C., Lorier, M., Carrell, R. W. 1978. *FEBS Lett.* 88:234–36
273. Shochat, D., Staples, S., Hargrove, K., Kozel, J. S., Chan, S. K. 1978. *J. Biol. Chem.* 253:5630–34
274. Johnson, D., Travis, J. 1979. *J. Biol. Chem.* 254:4022–26
275. Moroi, M., Aoki, N. 1976. *J. Biol. Chem.* 251:5956–65
276. Collen, D. 1976. *Eur. J. Biochem.* 69:209–16
277. Mullertz, S., Clemmensen, I. 1976. *Biochem. J.* 159:545–53
278. Wiman, B., Collen, D. 1979. See Ref. 21, pp. 145–56
279. Monkhouse, F. C., France, E. S., Seegers, W. H. 1955. *Circ. Res.* 3:397–402
280. Waugh, D. F., Fitzgerald, M. A. 1956. *Am. J. Physiol.* 184:627–39
281. Rosenberg, R. D., Damus, P. S. 1973. *J. Biol. Chem.* 248:6490–505
282. Markwardt, F., Walsmann, F. 1959. *Z. Physiol. Chem.* 317:64–77
283. Feinman, R. D. 1979. See Ref. 21, pp. 55–66
284. Petersen, T. E., Dudek-Wojciechowska, G., Sottrup-Jensen, L., Magnusson, S. 1979. See Ref. 21, pp. 43–54

J. Mol. Biol. (1998) 275, 347–363

JMB

# Kunitz-type Soybean Trypsin Inhibitor Revisited: Refined Structure of its Complex with Porcine Trypsin Reveals an Insight into the Interaction Between a Homologous Inhibitor from *Erythrina caffra* and Tissue-type Plasminogen Activator

**Hyun Kyu Song and Se Won Suh***

*Department of Chemistry and Center for Molecular Catalysis College of Natural Sciences Seoul National University Seoul 151-742, Korea*

*Corresponding author

The Kunitz-type trypsin inhibitor from soybean (STI) consists of 181 amino acid residues with two disulfide bridges. Its crystal structures have been determined in complex with porcine pancreatic trypsin in two crystal forms (an orthorhombic form at 1.75 Å resolution and a tetragonal form at 1.9 Å) and in the free state at 2.3 Å resolution. They have been refined to crystallographic *R*-values of 18.9%, 21.6% and 19.8%, respectively. The three models of STI reported here represent a significant improvement over the partial inhibitor structure in the complex, which was previously determined at a nominal resolution of 2.6 Å by the multiple isomorphous replacement method. This study provides the first high-resolution picture of the complex between a Kunitz-type proteinase inhibitor with its cognate proteinase. Many of the external loops of STI show high *B*-factors, both in the free and the complexed states, except the reactive site loop whose *B*-factors are dramatically reduced upon complexation. The reactive site loop of STI adopts a canonical conformation similar to those in other substrate-like inhibitors. The P1 carbonyl group displays no out-of-plane displacement and thus retains a nominal trigonal planar geometry. Modeling studies on the complex between a homologous Kunitz-type trypsin inhibitor DE-3 from *Erythrina caffra* and the human tissue-type plasminogen activator reveal a new insight into the specific interactions which could play a crucial role in their binding.



*Keywords:* crystal structure; Kunitz-type proteinase inhibitor; soybean trypsin inhibitor; β-trefoil fold; inhibition of plasminogen activator

Abbreviations used: BPT, bovine pancreatic trypsin; BPTI, bovine pancreatic trypsin inhibitor; ETI, trypsin inhibitor DE-3 from *Erythrina caffra*; ETIa, trypsin inhibitor a from *Erythrina variegata*; ETIb, trypsin inhibitor b from *Erythrina variegata*; Hepes, *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid; MCTI, trypsin inhibitor from bitter gourd seeds (*Momordica charantia*); MIR, multiple isomorphous replacement; ortho, orthorhombic crystal; PEG, polyethylene glycol; PPT, porcine pancreatic trypsin; PSTI, pancreatic secretory trypsin inhibitor; r.m.s., root-mean-square; SSI, *Streptomyces* subtilisin inhibitor; STI, Kunitz-type soybean trypsin inhibitor; tetra, tetragonal crystal; tPA, tissue-type plasminogen activator; WCI, Kunitz-type chymotrypsin inhibitor from winged bean seeds.

## Introduction

The Kunitz-type soybean trypsin inhibitor (STI), an archetypal member of its family, was characterized and crystallized as early as in 1947 (Kunitz, 1947a,b). Its partial structure in complex with porcine pancreatic trypsin (PPT) was reported in 1974 (Blow *et al.*, 1974; Sweet *et al.*, 1974), and amino acid sequences were reported for the three isoforms of STI, Ti$^a$, Ti$^b$, and Ti$^c$ (Kim *et al.*, 1985). They all contain 181 amino acid residues with two disulfide bridges (Cys39-Cys86 and Cys136-Cys145). Between the sequences of Ti$^a$ and Ti$^c$, there is only one substitution at position 55. On the other hand, eight amino acid substitutions exist between Ti$^a$ and Ti$^b$ (Kim *et al.*, 1985). Isolation of

0022-2836/98/020347-17 $25.00/0/mb971469

the cDNA clones encoding the precursors for $Ti^a$ and $Ti^b$ was reported (Song *et al.*, 1991).

The mechanism of inhibition of PPT by STI was proposed (Blow *et al.*, 1974; Baillargeon *et al.*, 1980). A stable complex between the virgin inhibitor and trypsin is rapidly formed and dissociates very slowly into the free enzyme and a mixture of unmodified and modified inhibitors. The modified STI is specifically cleaved at the scissile peptide bond between Arg63 (P1 residue) and Ile64 (P1' residue). In the refined structures of complex between bovine pancreatic trypsin inhibitor (BPTI) and bovine pancreatic trypsin (BPT), the scissile peptide bond remains nearly intact, with a slight out-of-plane deformation of the carbonyl oxygen atom being observed (Bode & Huber, 1992). However, the absence of a large upfield shift in the $^{13}C$ of the P1 residue of STI upon complex formation suggested that the STI:PPT complex is not a covalent, fully tetrahedral adduct (Baillargeon *et al.*, 1980). More recently, no out-of-plane distortion around the inhibitor's scissile peptide bond was observed in the refined structure of the complex between porcine β-trypsin and MCTI-A, a trypsin inhibitor from the squash family (Huang *et al.*, 1993).

Other Kunitz-type proteinase inhibitors from various sources show a high degree of homology to STI. Among them, three-dimensional structures have been reported for trypsin inhibitor DE-3 from *Erythrina caffra* (ETI) at 2.5 Å (Onesti *et al.*, 1991), proteinase K/α-amylase inhibitor from wheat at 2.5 Å (Zemike *et al.*, 1991), and Kunitz-type chymotrypsin inhibitor from winged bean seeds (WCI) at 2.95 Å (Dattagupta *et al.*, 1996), all in the free state. Other homologous proteins include a trypsin inhibitor from winged bean seeds (Yamamoto *et al.*, 1983), potato cathepsin D inhibitor (Mareš *et al.*, 1989), inhibitors of subtilisin/endogeneous α-amylase from barley and wheat (Leah & Mundy, 1989), and winged bean albumin-1 (Kortt *et al.*, 1989). Kunitz-type proteinase inhibitors are further divided into three groups based on their abilities to inhibit chymotrypsin, trypsin and tissue-type plasminogen activator (tPA). tPA is currently being used clinically as a thrombolytic agent. Group a inhibitors are relatively specific for chymotrypsin but are poor inhibitors of trypsin, and they do not inhibit tPA. Group b inhibitors are more specific towards trypsin than chymotrypsin and they do not inhibit tPA, either. Group c inhibitors inhibit trypsin, chymotrypsin, and tPA. Some proteinase inhibitors of the Kunitz-type in the seeds of *Erythrina* species belong to group c and are of particular interest due to their unique ability to bind to and to inhibit tPA (Heussen *et al.*, 1984). ETI, for example, provides an effective affinity reagent for the one-step purification of both one- and two-chain forms of tPA (Heussen *et al.*, 1984).

The crystal structure of the STI:PPT complex was previously determined at a nominal resolution of 2.6 Å by the MIR method (Blow *et al.*, 1974; Sweet *et al.*, 1974). However, only a very incomplete model of the inhibitor could be built, since the X-ray data used were relatively incomplete (56%) in the 3.0 to 2.6 Å range and a significant part of the inhibitor showed only poor electron density. Until now, there has been no report of a highly refined structure of any Kunitz-type proteinase inhibitor in complex with its cognate proteinase. Furthermore, the structure of STI in the free state has not yet been reported, precluding a detailed analysis of structural changes associated with complexation with a proteinase.

Therefore, there exists a need for high-resolution crystallographic analyses of STI, in complex with a proteinase as well as in the free state, in order to provide a high-resolution picture of STI and also to gain insight into the interactions between a Kunitz-type proteinase inhibitor with a serine proteinase such as trypsin. In this paper, we report the determination of three structures of STI, that is, the inhibitor alone at 2.3 Å resolution and its complex with PPT in two crystal forms, an orthothombic crystal form at 1.75 Å resolution and a tetragonal form at 1.9 Å. The structure of STI:PPT complex reported in this study provides a basis for modeling studies to gain insights into the interaction mode between homologous inhibitors and proteinases, including that between ETI and tPA. Modeling experiments were undertaken to understand why ETI, but not STI, is capable of inhibiting tPA.

## Results and Discussion

### Quality of refined models

Table 1 summarizes the refinement statistics as well as model quality parameters. All three models of STI:PPT complex in two crystal forms and free STI have reasonable crystallographic *R*-factors and excellent stereochemistry. In this paper, residues of PPT and other homologous proteinases are numbered according to the chymotrypsin sequence numbering scheme and a prime (') is appended to the residue number for clarity.

Missing from the inhibitor part of STI:PPT(ortho) model are two segments (residues 125 to 126 and 139 to 142) and the four C-terminal residues (178 to 181). In this model, the electron density for the side-chains of the following residues in the inhibitor, mostly in the loop regions, is either poor or absent: 98, 109, 111 and 112, 124, 127, 138, 143 and 144, 154 and 155, 165 to 167 and 177. The residues 125 and 126 missing in this model are clearly seen in the STI:PPT(tetra) model, whereas the segment 139to 143 and the four C-terminal residues (178 to 181) are still missing. Missing from the free STI model are two segments (125 and 126; 140 to 142) and the four C-terminal residues (178 to 181), with the electron density for the side-chains of the following residues being either poor or absent: 38, 98, 111 and 112, 124, 127, 138 and 139, 143 and 144, 154, 165 to 167, and 177. Only a single residue, Gln138, in the free STI model is in the disallowed region of Ramachandran plot. Electron density

**Table 1.** Refinement statistics

| | STI:PPT(ortho) | STI:PPT(tetra) | Free STI |
|---|---|---|---|
| Resolution range (Å) | 8.0–1.75 | 8.0–1.90 | 8.0–2.30 |
| No. of reflections ($F > 2\sigma$) | 31,038 | 23,967 | 9079 |
| Crystallographic *R*-factor (%) | 18.9 | 21.6 | 19.8 |
| Free *R*-factor (%)[a] | 21.4 | 28.0 | 26.9 |
| No. of residues | 171 (STI), 223 (PPT) | 172 (STI), 223 (PPT) | 172 |
| No. of non-hydrogen atoms | | | |
| Protein | 1268 (STI), 1630 (PPT) | 1286 (STI), 1630 (PPT) | 1,281 |
| Water | 142 | 139 | 44 |
| $Ca^{2+}$ | 1 (PPT) | 1 (PPT) | — |
| Average *B*-factor (Å²) | | | |
| Main-chain | 34.4 (STI), 19.6 (PPT) | 35.6 (STI), 19.1 (PPT) | 37.1 |
| Side-chain | 34.5 (STI), 23.8 (PPT) | 34.6 (STI), 20.6 (PPT) | 39.4 |
| Water | 37.2 | 34.2 | 36.0 |
| $Ca^{2+}$ | 19.6 (PPT) | 33.6 (PPT) | — |
| R.m.s. deviations from ideal geometry | | | |
| Bond lengths (Å) | 0.013 | 0.011 | 0.012 |
| Bond angles (°) | 1.32 | 1.68 | 1.40 |
| Ramachandran outliers | 0 | 0 | 1 (Gln138) |

[a] The free *R*-factor was calculated using 10% of the data.

around it is not well defined and its ψ angle lies slightly outside the allowed region. In the previously reported partial model of STI (Sweet *et al.*, 1974), residues 1 to 93 were built and refined with a certain degree of confidence but only a $C^{\alpha}$ tracing is available for the segments 94 to 106 and 130 to 176, while another stretch of density was tentatively assigned to residues 116 to 122. In the present models of STI, a highly flexible loop between 107 and 115 has been unambiguously traced.

Average *B*-factors for the main-chain atoms of STI in both the complexed and the free states are much higher than for PPT in the complex models (Table 1). This indicates a greater degree of inherent flexibility of the inhibitor. An average *B*-factor for the main-chain atoms of each residue in STI models is plotted as a function of residue number in Figure 1. Besides the C terminus for all three models and also the N terminus for the free inhibitor, regions with high *B*-factors correspond to long loops connecting β-strands. The electron densities for these regions with high *B*-factors are weak but the tracing of the polypeptide chain is not ambiguous. In all three models of STI, the region around Asn13 shows *B*-factors lower than the average (Figure 1). This is due to the involvement of this residue in the network of hydrogen bonds, which stabilizes the conformation of the reactive site loop. The final $(2F_o - F_c)$ electron density, calculated using data to 1.75 Å, for the reactive site loop of STI in the STI:PPT complex in the orthorhombic crystal is very well defined, reflecting the low *B*-factors around this region (Figure 2).

## Overall structure of STI

STI is roughly spherically shaped with approximate dimensions of 45 Å × 42 Å × 40 Å. The STI structure consists of twelve antiparallel β-strands, long loops connecting these β-strands, and a $3_{10}$-helix (Figure 3). It shows the same overall fold that has been described for the structures of ETI (Onesti *et al.*, 1991) and WCI (Dattagupta *et al.*, 1996). The reactive site (Arg63-Ile64) is located at one end of the structure on a protruding loop. Six of the strands (Aβ1, Aβ4, Bβ1, Bβ4, Cβ1, and Cβ4) form a short antiparallel β-barrel, with one side of the barrel being closed by a "lid" consisting of the other six strands. This common fold, termed the

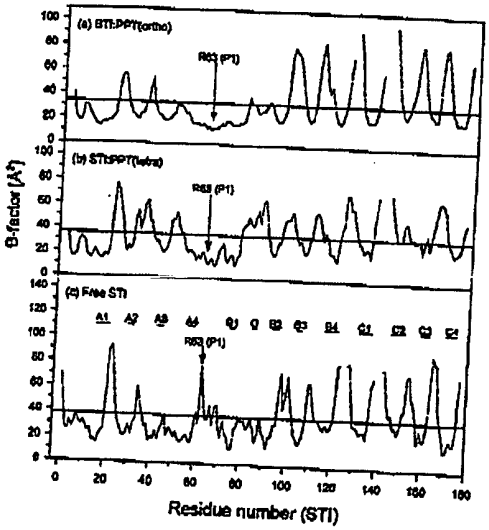


**Figure 1.** *B*-factor plot. Average *B*-factors for the main-chain atoms of each residue are plotted as a function of residue number for STI models (a) in STI:PPT complex in the orthorhombic crystal, (b) in STI:PPT complex in the tetragonal crystal; and (c) in the free state. Overall average values are indicated by horizontal lines. An assignment of secondary structure elements is also given (β-strands of subdomains A, B, and C: A1, residues 15 to 21; A2, 29 to 32; A3, 42 to 45; A4, 56 to 59; B1, 74 to 77; B2, 93 to 96; B3, 104 to 106; B4, 116 to 122; C1, 131 to 137; C2, 146 to 152; C3, 159 to 163; C4, 170 to 175; $3_{10}$-helix: G, 84 to 86).

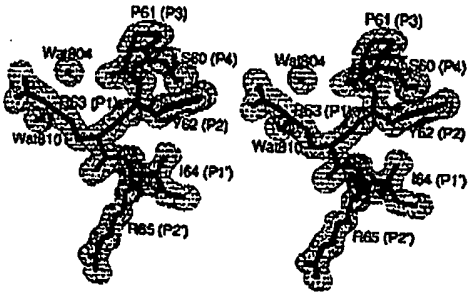


Figure 2. Final $(2F_o-F_c)$ electron density map around the reactive site loop (Ser60–Arg65; P4–P2') in the STI:PPT(ortho) structure. The map was calculated using 15.0 to 1.75 Å data and contoured at 1.2 σ.

"β-trefoil fold" (McLachlan, 1979; Murzin *et al.*, 1992), has also been observed in the structures of nonhomologous proteins including interleukin-1α (Graves *et al.*, 1990), interleukin-1β (Priestle *et al.*, 1988; Finzel *et al.*, 1989), ricin B-chain (Rutenber & Robertus, 1991), hisactophilin (Habazetti *et al.*, 1992), and fibroblast growth factors (Eriksson *et al.*, 1991; Zhang *et al.*, 1991; Zhu *et al.*, 1991).

The structure of STI displays an approximate 3-fold internal symmetry, with the symmetry axis coinciding with the barrel axis. The repeating unit is a four-stranded motif consisting of about 60 amino acid residues, structurally organized as L-β1-L-β2-L-β3-L-β4 (where L denotes a loop connecting consecutive β-strands). A superposition of the three subdomains shows high degree of similarity for the β-strands but not for the connecting loops (Figure 4). The r.m.s. deviations between topologically equivalent 16 $C^{\alpha}$ atoms (for residues 16 to 21, 28 to 32, 41 to 45 in subdomain A; residues 72 to 77, 92 to 96, 103 to 107 in subdomain B; residues 130 to 135, 147 to 151, 159 to 163 in subdomain C) are 1.0 Å between subdomains A and B, 1.1 Å between A and C, and 0.9 Å between B and C, respectively. In contrast to the presence of a $3_{10}$-helix at positions 84 to 86 in our STI models, ETI and WCI do not contain a $3_{10}$-helix in this region. This is due to three-residue deletions in ETI and WCI in this region. This $3_{10}$-helix in STI protrudes from the protein core (Figure 3) and the bond after Met84 is susceptible to cleavage by subtilisin (Laskowski *et al.*, 1974). Even when STI is cleaved after Met84, it is still active as a trypsin inhibitor, and when both peptide bonds after Arg63 (P1 residue) and Met84 are cleaved, the segment between Ile64 and Met84 does not dissociate, although it is held by non-covalent interactions (Laskowski *et al.*, 1974). This is understandable because the segment is an integral part of the six-stranded β-barrel. The two disulfide bridges (Cys39-Cys86 and Cys136-Cys145) in STI are well separated from each other in the structure (Figure 3(a)). These disulfide bonds apparently

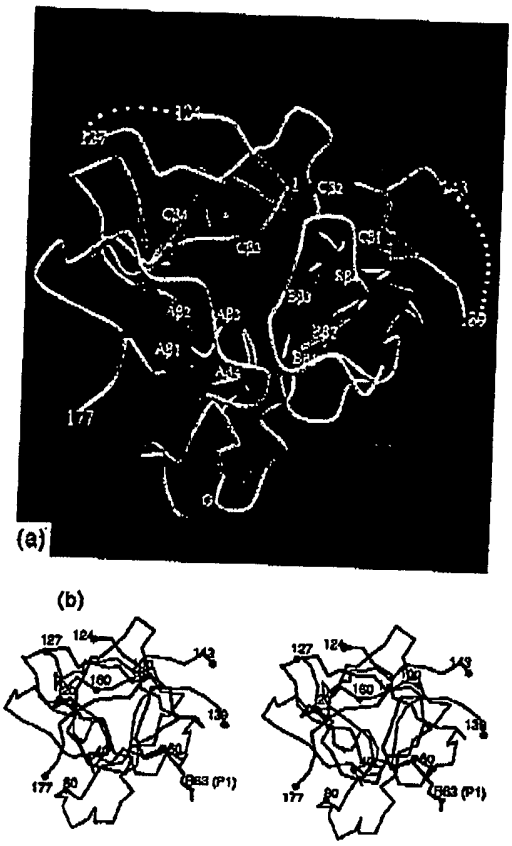


Figure 3. Overall fold of STI. (a) Ribbon diagram of STI. Green arrows, a red ribbon and light pink ropes represent β-strands, a $3_{10}$-helix and the loops, respectively. The P1 residue Arg63 is indicated as well as the two disulfide bonds (C39–C86 and C136–C145). The secondary structure elements are also labeled. (b) A stereo view of the $C^{\alpha}$ trace of STI. The P1 residue Arg63 is drawn and approximately every twentieth residue is marked by a dot.

reduce the flexibility of the loop regions by cross-linking them.

## Solvent structure

In the early structure of STI in its complex with PPT (Sweet *et al.*, 1974), no solvent structure was reported. In this study, all three structures of STI in its complex with PPT and in its free state have been refined at high enough resolutions so that ordered water molecules can be located with confidence. The refined model of free STI includes a total of 44 ordered solvent molecules, all modeled as water. Most of them are bound to the protein surface. Two water molecules (Wat804 and Wat810) are involved in mediating the interaction between STI and PPT in the complex. They occupy equivalent positions in both orthorhombic and

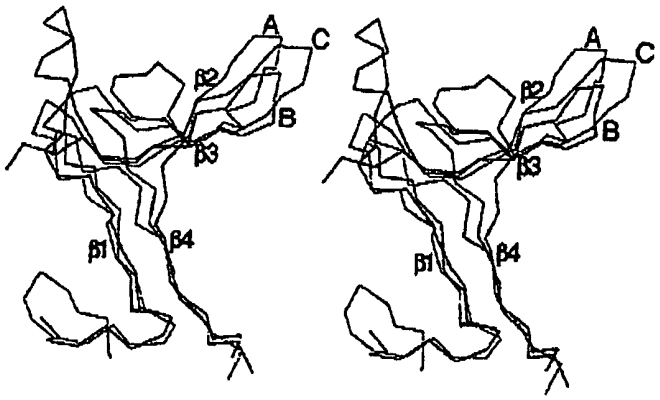


Figure 4. Stereo diagram showing the superposition of $C^{\alpha}$ atoms among three subdomains (A, B, and C) in STI structure. Red, blue and green lines represent domains A, B and C, respectively. The numbers (1 to 4) indicate the four β-strands in each subdomain.

tetragonal crystal forms and show B-factors lower than the average (13.4 and 16.5 Å$^2$, respectively). The water molecule Wat804 in the active-site pocket occupies a position to form hydrogen bonds with some of Arg63 NH2, Pro61 O, Gln192′ OE1, Gly215′ O, and Gly219′ O atoms. Another water molecule, Wat810, makes hydrogen bonds with Arg63 NH1 and Ser190′ OG atoms. Water molecules equivalent to Wat804 and Wat810 of STI:PPT complex are also found in other trypsin-inhibitor complexes: Wat403 and Wat416 in BPTI:BPT (Marquart *et al.*, 1983) and Wat58 and Wat5 in MCTI:PPT (Huang *et al.*, 1993). An interesting integral water molecule is present in the center of a triangular structure made up of β-strands Aβ3, Bβ3 and Cβ3, located almost on the pseudo three-fold axis (Wat34 in Figure 5). An equivalent water molecule is also present in the ETI model, but a water molecule which mediates the interaction among the three β-strands at the bottom of β-barrel in the ETI model (Wat220 in Onesti *et al.*, 1991) is not observed in the STI models.

### Comparison of inhibitor structures

When the refined structure of STI in the STI:PPT(ortho) complex model is superposed with the $C^{\alpha}$ coordinates of STI in the early structure of the complex (Sweet *et al.*, 1974), which were kindly supplied to us by Drs David Blow and Silvia Onesti during the final stage of our manuscript preparation, an r.m.s. deviation of 1.98 Å is calculated for 147 $C^{\alpha}$ atom pairs (residues 1 to 97, 116 to 122, 130 to 138, and 143 to 176), with the maximum $C^{\alpha}$ difference of 6.02 Å. Since the early STI model has not been refined well and is rather incomplete, our detailed structural comparison is limited to the refined models of STI in the free state and in its complex with PPT, which are reported in this study. A summary of the superpositions among STI models is given in Table 2, which lists the rms differences for $C^{\alpha}$ atom pairs. Table 2 also gives the results for comparing the trypsin models as well as for comparing the structures of three homologous inhibitors in their free state. It can be seen from Table 2 that the structural variability is greater for STI than for PPT.

### *Comparison of the free and trypsin-complexed STI models*

All three STI models are superposed in Figure 6(a). The most striking change in STI upon complexation with PPT is a lowering in *B*-factors of its reactive site loop (residues 60 to 65) and also of the N terminus (Figure 1). This indicates that the

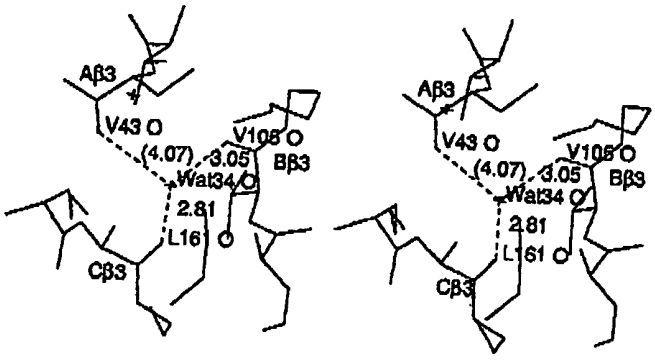


Figure 5. Stereo diagram showing the interactions between the bound water molecule (Wat34) and three β-strands (Aβ3, Bβ3 and Cβ3). The distances between the water and oxygen atoms of the inhibitor are given, with a distance too long for a hydrogen bond in parenthesis.

Table 2. R.m.s. difference for $C^{\alpha}$ superposition among the trypsin and inhibitor models (Å)

| PPT | STI | STI family inhibitors |
|---|---|---|
| MCT[a]–PPT(ortho), 0.29 (223)[b] | Free STI–STI(ortho), 0.68 (171) | Free STI–ETI, 2.39 (160) |
| MCT–PPT(tetra), 0.35 (223) | Free STI–STI(tetra), 0.79 (170) | Free STI–WCI, 2.37 (165) |
| PPT(ortho)–PPT(tetra), 0.30 (223) | STI(ortho)–STI(tetra), 0.50 (170) | WCI–ETI, 1.10 (158) |

[a] Starting porcine β-trypsin model for molecular replacement.
[b] The value in the parenthesis is the number of $C^{\alpha}$ pairs for calculating the r.m.s. difference.

relatively flexible reactive site loop as well as the N terminus of STI becomes rigid upon its binding to trypsin. Similar behaviors have been previously observed for the reactive site loops of other inhibitors (Read & James, 1986).

A $C^{\alpha}$ superposition of STI models in the free state and its STI:PPT(ortho) complex gives an r.m.s. difference of 0.68 Å for 171 $C^{\alpha}$ atom pairs (1.09 Å for all 1262 non-hydrogen atoms). A plot of the distance between equivalent $C^{\alpha}$ atoms as a function of residue number is given in Figure 7(a). Residues showing deviations greater than 1.0 Å for $C^{\alpha}$ atoms are 22 and 23, 98, 111, 124, 128, 138, 144, 154 and 155, 166, and 177. A $C^{\alpha}$ superposition of the free STI and STI:PPT(tetra) models gives an r.m.s. difference of 0.79 Å for 170 $C^{\alpha}$ atom pairs (1.23 Å for all 1248 non-hydrogen atoms). A plot of the distance between equivalent $C^{\alpha}$ atoms as a function of residue number is given in Figure 7(b). Residues showing deviations greater than 1.0 Å for $C^{\alpha}$ atoms are 48 to 49, 124, 144, 155, 166 and 167, and 177. In both cases, the large differences are confined to the external loops and the C-terminal residue Asp177. These regions show high *B*-factors for the main-chain atoms (Figure 1), indicating that the coordinate errors of these regions are larger than the average. Thus large deviations in these regions are partly due to the large coordinate errors but they also reflect the inherent conformational variability of these regions. In other parts of the chain, the magnitude of the deviation is roughly what is expected from the estimated coordinate errors. A superposition of the STI models in the two complex structures gives an r.m.s. difference of 0.50 Å for 170 $C^{\alpha}$ atom pairs (0.75 Å for all 1264 non-hydrogen atoms). A plot of the distance between equivalent $C^{\alpha}$ atoms as a function of residue number is given in Figure 7(c). Residues showing deviations greater than 1.0 Å for $C^{\alpha}$ atoms are 48 and 49, 111, 128, 155, 167, and 177. This large

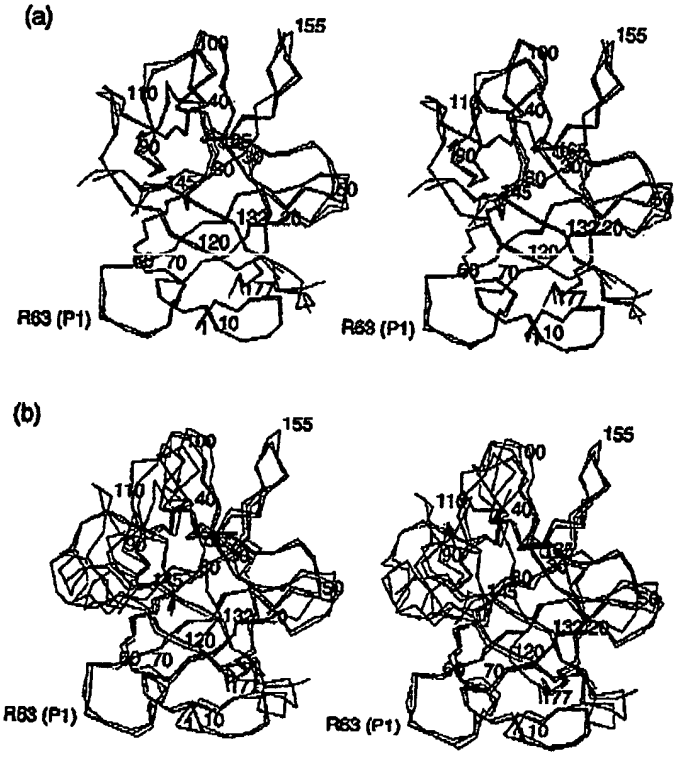


Figure 6. (a) Stereo diagram showing the superposition of $C^{\alpha}$ atoms among the free and complexed STI models. Green lines represent the free STI, while red and blue lines represent the inhibitor in STI:PPT(ortho) and STI:PPT(tetra) models, respectively. (b) Stereo diagram showing the superposition of $C^{\alpha}$ atoms among the STI family inhibitors. Green, red and blue lines represent STI, ETI and WCI, respectively. Every twentieth residue is labeled.

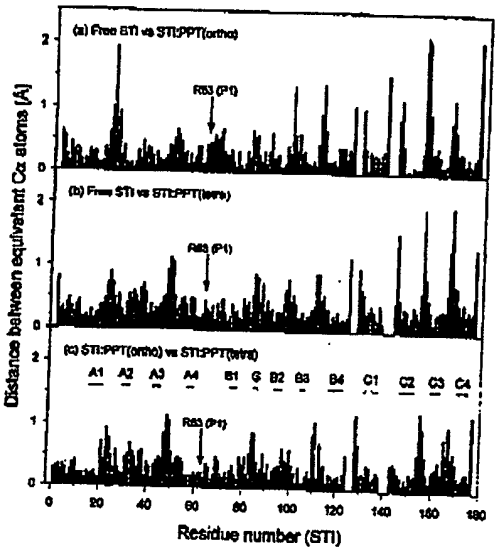


Figure 7. Plot of the distance between equivalent $C^{\alpha}$ atoms between the free and complexed STI models as a function of residue number. (a) Free STI *versus* STI:PPT(ortho), (b) Free STI *versus* STI:PPT(tetra); and (c) STI:PPT(ortho) *versus* STI:PPT(tetra).

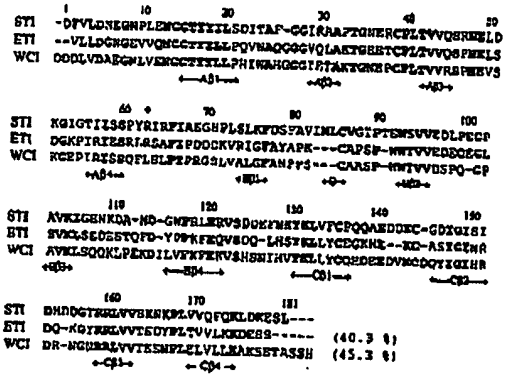


Figure 8. Sequence alignment of STI, ETI and WCI. The numbering is for STI. The secondary structure elements are assigned as in Figure 1. Residues conserved in all three sequences are in bold and the bullet (•) denotes the P1 residue.

discrepancy has contributions from the large coordinate errors in these regions as well as from the different loop conformations due to different crystalline environments, as in the comparison between the free and complexed STI models. Therefore, the structural difference between the two models of STI in different environments of the orthorhombic and tetragonal crystals of the complex should be regarded as only minor. In particular, an r.m.s. difference for the seven $C^{\alpha}$ atoms in the reactive site loop (P4–P3′) is only 0.17 Å. Thus, the model of STI:PPT(ortho), which has been refined with a more complete data extending to 1.75 Å resolution, will be used for most of the structure descriptions below, unless otherwise stated.

### *Comparison of free STI with ETI and WCI*

The alignment of amino acid sequences of STI, ETI, and WCI based on their three-dimensional structures is shown in Figure 8. The sequence identity between ETI and WCI is 57%, whereas that between ETI and STI sequences is 40% and that between WCI and STI is 45%. This is also manifested in the results of structural comparisons. Figure 6(b) shows a $C^{\alpha}$ superposition of the models of free STI, ETI, and WCI, refined at 2.3, 2.5, and 2.95 Å, respectively. Between the free STI and ETI, the r.m.s. deviation is 2.39 Å for 160 $C^{\alpha}$ atom pairs, with the residues 23, 63, 82 to 84, 88, 108, 110 to 116, 138, and 167 of STI showing $C^{\alpha}$ deviations greater than 3 Å. Between the free STI and WCI, the r.m.s. deviation is 2.37 Å for 165 $C^{\alpha}$ atom pairs, with the residues 82 to 84, 86, 88 and 89, 98 and 99, 110 to 116, 124, 127, and 143 of STI showing $C^{\alpha}$ deviations greater than 3 Å. When ETI and WCI models are compared, the r.m.s. deviation is 1.10 Å for 158 $C^{\alpha}$ atom pairs, only about half as large as in comparing STI with either ETI or WCI (Table 2). A plot of the distance between equivalent $C^{\alpha}$ atoms as a function of residue number is given in Figure 9. Apparently, the difficulty in map interpretation of free STI (especially, for residues 80 to 88 and 110 to 115) using the phases from the starting ETI model only is due to the large (>6 Å) difference between the structures of ETI and STI (Figure 9). Other discrepant regions are located mostly on the protruding loops of the inhibitors, where insertions or deletions occur.

In the case of ETI, the reactive site loop is displaced with respect to that of free STI, with the distance of 3.15 Å between the $C^{\alpha}$ atoms of P1 residues. The corresponding distance between ETI and STI:PPT(ortho) is 2.78 Å. The reactive site loop of WCI lies between those of STI and ETI (Figure 6(b)) and its displacement is much smaller, with the distance of 1.02 Å between the $C^{\alpha}$ atoms of P1 residues in free STI and WCI. Thus the large displacement of the reactive site loop of ETI is largely due to the crystal packing, but is not a consequence of the difference in the complexation state, as speculated by the authors (Onesti *et al.*, 1991). This argument is further supported by an excellent correlation between the $C^{\alpha}$ distance plot for comparing the complexed STI with ETI and that for comparing the free STI and ETI (Figure 9(a) and (b)). Clearly, most of the discrepancy between the complexed STI and free ETI, including the displacement of the reactive site loop, cannot be ascribed to the difference in

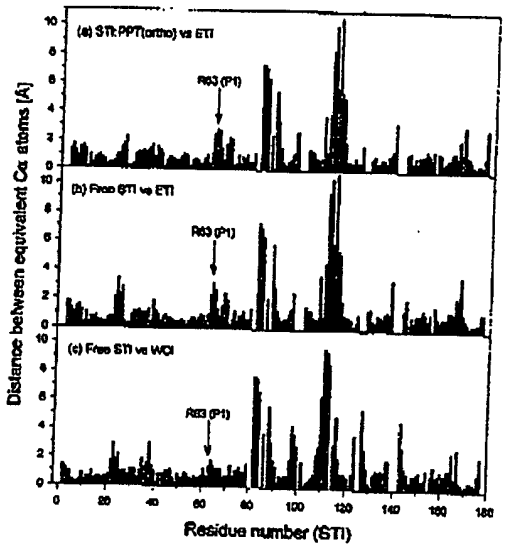


Figure 9. Plot of the distance between equivalent $C^{\alpha}$ atoms between the STI family inhibitors as a function of residue number. (a) STI:PPT(ortho) *versus* ETI, (b) Free STI *versus* ETI, and (c) Free STI *versus* WCI.

the complexation state. When $C^{\alpha}$ atoms for P4–P3′ residues are superposed, the r.m.s. deviation is 0.71 Å between the free STI and ETI, while it is 0.56 Å between the free STI and WCI. The corresponding r.m.s. deviations are 0.36 Å between the free STI and STI:PPT(ortho) and 0.49 Å between the free STI and STI:PPT(tetra). The local conformations of reactive site loops of STI, ETI, and WCI are, therefore, very similar and are little affected upon complex formation.

## The reactive site loop: its conformation and changes upon complexation

STI belongs to the family of substrate-like inhibitors, which possess an exposed reactive site loop of a characteristic canonical conformation. The reactive site loop of STI is not constrained by the secondary structure elements or disulfide bridges that could limit its conformational freedom, unlike many other proteinase inhibitors. In STI, the side-chain of Asn13 plays an important role in stabilizing the reactive loop conformation through a network of hydrogen bonds (Figure 10). This results in the low *B*-factors for the region around Asn13 (Figure 1). Similar stabilizations are also observed for ETI and WCI, although the details of hydrogen bonding pattern are not identical. For example, a water molecule resides in the reactive site loop of WCI, forming hydrogen bonds with the carbonyl O atoms of Phe64 and Ser62 (Dattagupta *et al.*, 1996). An equivalent water molecule is not observed in either the STI or the ETI structure. The backbone conformation of the reactive site loop of STI does not change significantly upon forming a complex with PPT, as indicated by similar main-chain dihedral angles for P4–P3′ residues (Table 3). This is similarly found in BPTI and its complex with BPT (Table 3). Table 3 shows that there is a large deviation of $\phi$ and $\psi$ angles at several positions between our refined models of STI and the early model of STI in its complex with PPT (Sweet *et al.*, 1974). Table 3 also lists the conformational angles of several other inhibitors as well as their amino acid sequences around the scissile peptide bond. Between STI and BPTI, there is a large deviation in the $\phi$ angle of P4 and $\phi$ and $\psi$ angles of P2′. Despite a great difference in the main-chain conformational angles at P2′ position (arginine) for STI and BPTI, both guanidinium groups point in the same direction as pointed out earlier (Sweet *et al.*, 1974). Between STI and MCTI, a large deviation is observed for $\phi$ and $\psi$ angles of P3 and the $\psi$ angle of P2′, whereas ETI and WCI show similar conformations as STI at P2′. Notwithstanding some differences in the main-chain conformations at some positions, the side-chains, which interact with the proteinase, are quite similarly oriented in several substrate-like inhibitors.

The extension of the reactive site loop of STI family (measured as the distance between $C^{\alpha}$

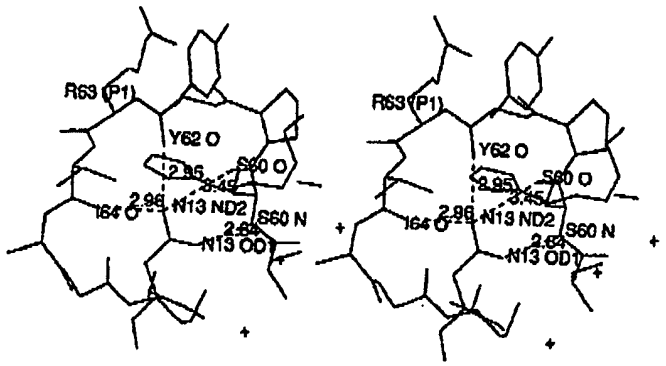


Figure 10. Stereo diagram showing the role of Asn13 in stabilizing the reactive site loop conformation of STI. The conserved residue Asn13 is involved in extensive hydrogen bondings with several residues in the reactive site loop.

Table 3. Comparison of dihedral angles ($\phi/\psi$) of reactive site loop residues

| | P4 | P3 | P2 | P1 | P1′ | P2′ | P3′ |
|---|---|---|---|---|---|---|---|
| STI | S | P | Y | R | I | R | F |
| ETI | S | R | L | R | S | A | F |
| WCI | S | Q | F | L | S | L | F |
| BPTI | G | P | C | K | A | R | I |
| MCTI | I | C | P | R | I | W | M |
| Free STI | −112/149 | −70/−28 | −62/158 | −84/21 | −60/148 | −90/−22 | −127/126 |
| STI:PPT(ortho) | −114/144 | −58/−34 | −56/139 | −89/38 | −83/148 | −67/−38 | −118/155 |
| STI:PPT(tetra) | −126/149 | −67/−23 | −65/121 | −70/10 | −59/165 | −88/−29 | −115/170 |
| STI:PPT(Blow) | −58/127 | −40/−39 | −65/141 | −93/47 | −93/123 | −165/−178 | 8/49 |
| ETI | −88/171 | −66/−30 | −112/150 | −66/134 | −169/151 | −74/−46 | −106/163 |
| WCI | −125/−177 | −85/−21 | −80/173 | −111/76 | −122/158 | −89/−19 | −130/157 |
| BPTI | 86/176 | −86/−6 | −81/165 | −104/9 | −76/172 | −127/76 | −105/121 |
| BPTI:BPT | 78/174 | −77/−29 | −70/155 | −116/39 | −87/164 | −112/79 | −98/124 |
| MCTI | −58/117 | −137/112 | −54/147 | −91/34 | −82/146 | −84/125 | −152/127 |

STI:PPT(Blow), Sweet *et al.* (1974); ETI, Onesti *et al.* (1991); WCI, Dattagupta *et al.* (1996); BPTI, Wlodawer *et al.* (1987); BPTI:BPT, Marquart *et al.* (1983); MCTI, Huang *et al.* (1993).

atoms of residues P4 and P3′) is among the smallest (9.0 Å for STI, 9.2 Å for ETI, and 8.8 Å for WCI compared with 11.3 Å for BPTI, 15.0 Å for MCTI (squash family), 15.6 Å for PSTI (Kazal family), and 15.9 Å for eglin c (PI-1 family)). The reactive site loops of BPTI (Kunitz), Kazal, squash, and PI-1 families are constrained by disulfide bridges that could limit their conformational freedom. No disulfide bridge adjacent to the reactive site is present in the STI and eglin c (PI-1 family). Whereas the reactive site loop of STI is held in a favorable conformation by hydrogen bonds (Figure 10), that of eglin c is stabilized by electrostatic interactions between charged residues on both sides of the reactive site loop.

The magnitude of overall structural changes in proteinase inhibitors upon binding the cognate proteinases ranges from very small to extremely large. For the best-characterized BPTI, the r.m.s. difference between the free and complexed structures is 0.26 to 0.27 Å for 58 $C^{\alpha}$ atom pairs (Perona *et al.*, 1993). The r.m.s. difference between the free *Streptomyces* subtilisin inhibitor (SSI) and the inhibitor in SSI:subtilisin BPN′ complex is 0.63 Å for 107 $C^{\alpha}$ atom pairs (Takeuchi *et al.*, 1991). At the other extreme, the serpin family shows the most dramatic conformational change. The structure of cleaved $\alpha_1$-antitrypsin showed an unexpected separation of its reactive center peptide (Met358-Ser359; P1–P1′) by about 70 Å with incorporation of the reactive loop into the β-sheet structure (Loebermann *et al.*, 1984). A superposition of 54 $C^{\alpha}$ atoms in the central β-sheet of the cleaved and uncleaved $\alpha_1$-antitrypsin gave an r.m.s. difference of 2.4 Å (Song *et al.*, 1995). For STI, the r.m.s. difference between the free and complexed inhibitors is 0.68 to 0.79 Å for 171 $C^{\alpha}$ atom pairs. The deviation for $C^{\alpha}$ atoms belonging to P4–P3′ residues in the reactive site loop is smaller than the overall change: 0.40 Å between the free STI and STI:PPT(ortho) and 0.26 Å between the free STI and STI:PPT(tetra).

## Geometry of the reactive site carbonyl group

The geometry of the carbonyl group at P1 residue is of great importance in understanding the interaction between the inhibitor and the proteinase during the catalytic mechanism. In the earlier and lower resolution structure of STI:PPT complex (Sweet *et al.*, 1974), a tetrahedral intermediate was proposed. This proposal was not, however, validated by the subsequent high-resolution structures of other complexes, because only slight (Marquart *et al.*, 1983; Bode & Huber, 1992) or negligible (Read *et al.*, 1983; Huang *et al.*, 1993) out-of-plane deformations of the P1 carbonyl group were observed. Additionally, the carbonyl carbon atom was found to be within the "sub-van-der-Waals" distance (typically around 2.7 Å) from the Ser195′ OG atom (Bode & Huber, 1992). In our STI:PPT(ortho) and STI:PPT(tetra) models, the distance is 2.76 and 2.71 Å, respectively. In order to obtain an unbiased geometry of the P1 carbonyl group, refinements were carried out either with or without applying the restraints to the carbonyl carbon atom. With the usual restraints, the out-of-plane distance of Arg63 C is 0.0084 and 0.0156 Å for STI:PPT(ortho) and STI:PPT(tetra), respectively (the positive value corresponding to a distortion toward the Ser195′ OG of trypsin). Without the restraints, it increases to 0.024 and 0.067 Å for STI:PPT(ortho) and STI:PPT(tetra), respectively. As a control experiment, refinements were also performed after removing the restraints for the carbonyl carbon atom of P2′ residue (Arg65). The out-of-plane distance of Arg65 C shows a considerable increase from 0.019 to 0.082 Å for STI:PPT(ortho) or from 0.026 to 0.093 Å for STI:PPT(tetra), respectively, when the restraints on the carbonyl carbon atom of Arg65 are removed. Therefore, it is concluded that the P1 carbonyl group of STI displays no significant out-of-plane displacement and thus it retains a nominal trigonal planar when complexed with PPT. This is in agreement with the

result for the MCTI:PPT complex (Huang *et al.*, 1993) and is consistent with STI belonging to a typical substrate-like inhibitor family, whose reactive site loop takes the canonical conformation in both the free and the complexed states. It is worth mentioning that the out-of-plane distance of Lys15 (P1) C ranges between 0.12 and 0.14 Å for BPTI complexes (PDB ID codes: 2PTC, 1TPA, and 1TGP).

### Mode of interaction between STI and PPT

The pattern of interaction between STI and PPT is depicted in Figure 11(a) and is summarized in Table 4. There are some minor differences between the interaction pattern between the orthorhombic and tetragonal crystal structures of the complex. Twelve amino acid residues out of the 181 in STI make contact with PPT in the orthorhombic crystal structure. They are: Asp1, Phe2, Asn13, Pro61 (P3), Tyr62 (P2), Arg63 (P1), Ile64 (P1′), Arg65 (P2′), His71, Pro72, Trp117 and Arg119. In the tetragonal crystal structure, the three residues His71, Pro72 and Arg119 do not interact with PPT. However, the pattern for the hydrogen bonding interaction involving the reactive site loop residues from Pro61 (P3) to Arg65 (P2′) is well conserved between the two crystal forms, except that a hydrogen bond between Tyr62 and Gly96′ is not present in the tetragonal crystal structure (Table 4).

Our orthorhombic crystal of the STI:PPT complex, although grown under a condition different from that of Sweet *et al.* (1974), has the same space group with nearly identical cell parameters. However, the interaction pattern observed in our orthorhombic structure is somewhat different from that described by Sweet *et al.* (1974). We believe that this difference is mainly due to the fact that our model is more complete and is better refined at higher resolution. The protein sequencing data (Kim *et al.*, 1985) and the nucleotide sequence of cDNA (Song *et al.*, 1991) indicate that the amino acid residues at P4 and P3 positions in the early model of the complex, Pro60 (P4) and Ser61 (P3), are erroneous. The corrected sequence, Ser60 (P4) and Pro61 (P3), was built into our model and is confirmed by the excellent electron density (Figure 2). The residues Pro60 (P4) and Phe66 (P3′) in the early model, suggested to form the region of interaction with PPT, are not involved in the interaction in both of our models. The two residues Trp117 and Arg119 belong to the segment 116 to 122, which was tentatively assigned in the early model of the complex and were not included in the residues participating in the interaction with trypsin (Sweet *et al.*, 1974). A detailed description of more differences is given below.

Most of the contacts between STI and PPT involve the five residues in the reactive site loop (P3-P2′). In our orthorhombic crystal structure, eleven of the 14 hydrogen bonds between STI and PPT are accounted for by the above residues, while in the tetragonal structure, ten of the eleven are accounted for. The P1 residue Arg63 makes the

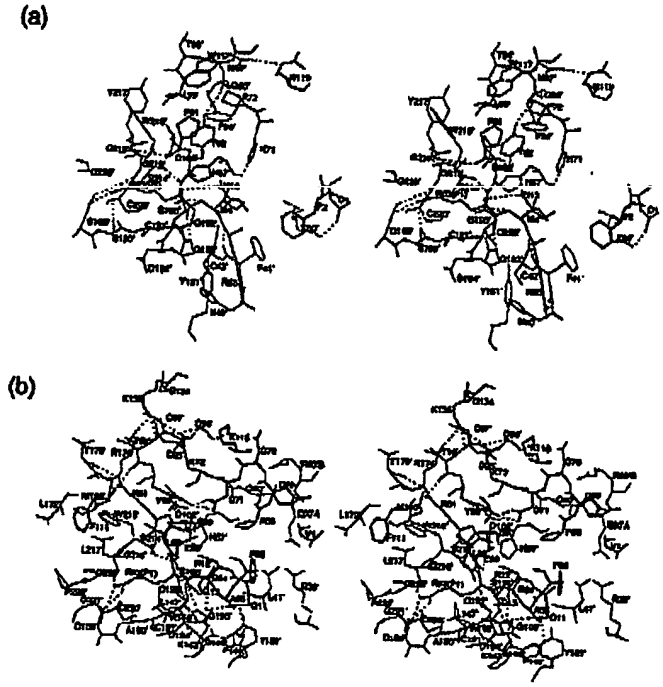


Figure 11. Stereo diagram showing the interaction; (a) between STI and PPT; and (b) between BTI and tPA. Blue and red lines represent the inhibitor and the proteinase, respectively. All residues involved in the interaction are labeled. Dotted lines denote hydrogen bonds.

Table 4. Total interactions between STI and PPT

| STI | PPT | [a]No. of interactions STI:PPT(ortho) | STI:PPT(tetra) | [b]Note |
|---|---|---|---|---|
| Asp1 | Lys60′ | 2, — | —, *1* (OD2–NZ; 2.80 Å) | ion |
| Phe2 | Phe41′ | 3, — | 5, — | a–a |
| | Lys60′ | 3, — | 3, — | |
| Asn13 | Gln192′ | *4, 1* (ND2–OE1; 3.46 Å) | 5, — | |
| Pro61 | Trp215′ | 3, — | 3, — | |
| | Gly216′ | *1, 1* (O–N; 3.12 Å) | *4, 1* (O–N; 3.14 Å) | |
| Tyr62 | His57′ | 16, — | 5, — | |
| | Phe94′ | 2, — | —, — | |
| | Gly96′ | —, *1* (OH–O; 2.97 Å) | —, — | |
| | Leu99′ | 5, — | 8, — | |
| | Asp102′ | 1, — | —, — | |
| | Gln192′ | 1, — | —, — | |
| | Ser214′ | 1, — | 1, — | |
| | Trp215′ | 1, — | 1, — | |
| Arg63 | Asp189′ | *4, 2* (NH1–OD1; 2.70 Å) (NH2–OD2; 3.09 Å) | *4, 2* (NH1–OD1; 2.68 Å) (NH2–OD2; 2.95 Å) | ion |
| | Ser190′ | *4, 1* (NH1–OG; 3.11 Å) | *3, 1* (NH1–OG; 3.43 Å) | ion |
| | Cys191′ | 4, — | 5, — | |
| | Gln192′ | 4, — | 4, — | |
| | Gly193′ | *2, 1* (O–N; 2.61 Å) | *2, 1* (O–N; 2.54 Å) | |
| | Asp194′ | 1, — | —, — | |
| | Ser195′ | 8, — | 5, — | |
| | Ser214′ | *1, 1* (N–O; 3.03 Å) | —, *1* (N–O; 3.38 Å) | |
| | Trp215′ | 2, — | 1, — | |
| | Gly216′ | 1, — | —, — | |
| | Gly219′ | —, *1* (NH2–O; 2.79 Å) | —, *1* (NH2–O; 2.82 Å) | |
| | Cys220′ | —, — | 3, — | |
| | Gly226′ | 1, — | —, — | |
| Ile64 | Phe41′ | —, — | 1, — | |
| | Cys42′ | 2, — | 1, — | |
| | His57′ | 1, — | 3, — | |
| | Gln192′ | 1, — | 2, — | |
| | Gly193′ | 1, — | 1, — | |
| | Ser195′ | *2, 1* (N–OG; 3.12 Å) | *1, 1* (N–OG; 2.89 Å) | |
| Arg65 | His40′ | *1, 1* (NH2–O; 3.07 Å) | *1, 1* (NH2–O; 2.99 Å) | |
| | Phe41′ | —, *1* (N–O; 3.34 Å) | —, *1* (N–O; 3.41 Å) | |
| | Tyr151′ | 3, — | 6, — | a–n |
| | Gly193′ | —, — | 2, — | |
| His71 | His57′ | *1, 1* (NE2–O; 3.08 Å) | —, — | |
| Pro72 | Asn97′ | 1, — | —, — | |
| | Leu99′ | 1, — | —, — | |
| Trp117 | Asn97′ | 2, — | —, — | |
| | Tyr217′ | —, — | 5, — | |
| Arg119 | Asn97′ | *1, 1* (NE–OD1; 3.21 Å) | —, — | |

[a] An interatomic contact is counted for every pair of atoms which is within 0.5 Å of theoretical van der Waals contact distance. The table indicates the number of interatomic contacts and hydrogen bonds (in *italics*) between amino acids of STI and those of PPT.
[b] Ion: possible ionic interaction, a–a: aromatic–aromatic edge–face interaction, a–n: aromatic–NH group interaction.

most extensive hydrogen bonds with PPT, forming six hydrogen bonds in total. The side-chain of Arg63 occupies its expected position in the primary binding pocket of PPT. The guanidium group of Arg63 makes an ionic interaction with the carboxylate group of Asp189′ in PPT. Hydrogen bonds are possible between the nitrogen atoms of the side-chain of Arg63 and the side-chain oxygen atom (OG) of Ser190′ and the carbonyl O atom of Gly219′. This differs from the hydrogen bonding pattern reported previously (Sweet *et al.*, 1974). The phenolic side-chain of Tyr62 (P2) is positioned between the side-chains of Leu99′ and His57′, lying parallel to the imidazole ring of the latter. Its hydroxyl group forms a hydrogen bond with the carbonyl O atom of Gly96′. It was suggested to form a hydrogen bond with the side-chain of Asn97′ in the previous report (Sweet *et al.*, 1974). Even though the residue Pro61 (P3) was erroneously assigned as Ser in the early model, the main-chain dihedral angles at this position (Table 3) and the hydrogen bonding interaction between its carbonyl O atom and the main-chain N atom of Gly216′ are similar. The long side-chain of Arg65 (P2′) interacts with Tyr151′ in PPT, possibly forming a hydrogen bond between its nitrogen NE atom and the aromatic ring (Levitt & Perutz, 1988).

Gln192′ of PPT belongs to a stretch of eight residues strictly conserved between PPT and tPA (Figure 12) and interacts with as many as four residues of STI (Asn13, Tyr62, Arg63 and Ile64). Therefore, its conformation is of great importance in the prediction of STI:tPA complex structure by homology modeling. Its unique conformation could not

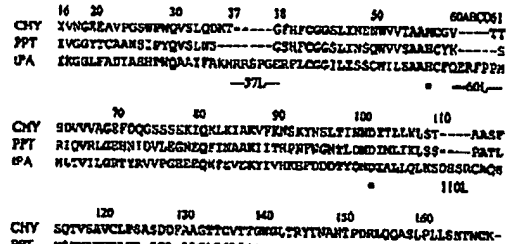

Figure 12. Sequence alignment of serine proteinases based on their three-dimensional structures. The abbreviations are as follows. Bovine chymotrypsin (CHY), porcine pancreatic trypsin (PPT), and human tissue-type plasminogen activation (tPA). The numbering scheme is according to the sequence number of CHY. The bullets (•) denote the positions of catalytic residues His57, Asp102 and Ser195. Several loop regions (L) of tPA are indicated.

be assigned in the early model of Sweet *et al.* (1974), because the electron density for the side-chain was not long enough and there was little room for the side-chain in the complex. In contrast, the side-chain of Gln192′ in PPT is well defined by the electron density in our models. This residue was suggested to form a hydrogen bond with the P4 residue of STI (Sweet *et al.*, 1974), but this interaction is absent in both of our complex models.

A close contact between the aromatic rings of Phe2 in STI and Phe41′ in PPT was noted earlier (Sweet *et al.*, 1974), with 13 interatomic contacts. In our structure, the two aromatic rings make an edge–face interaction, resulting in much fewer interatomic contacts. This kind of aromatic edge–face interaction is energetically favorable and is the most common among the observed aromatic–aromatic interactions (Burley & Petsko, 1985). A similar interaction was found in the crystal structure of chymotrypsin complexed with 6-benzyl-3-chloro-2-pyrone (Ringe *et al.*, 1985). The ionic interaction between Asp1 and Lys60′ was also noted earlier (Sweet *et al.*, 1974) but the side-chains were drawn unreasonably close together by the real-space refinement procedure. In our models, they are much better defined and there is no extensive van der Waals contact between them.

### A structural basis of tPA inhibition by ETI

There is a strong clinical interest in human tPA (Collen & Lijnen, 1995). Figure 12 compares the amino acid sequences of chymotrypsin, PPT and tPA, as aligned on the basis of their three-dimensional structures. In this paper, residue numbers of all serine proteinases follow the chymotrypsin numbering for consistency. Human tPA shows a sequence identity of 41% against PPT. The recent crystal structure of recombinant two-chain human tPA showed that the S1 pocket of tPA is almost identical to that of trypsin, whereas the S2 site is considerably reduced in size (Lamba *et al.*, 1996; PDB ID code, 1RTF).

Although the three inhibitors STI, ETI and WCI show a high degree of sequence identity (Figure 8), only ETI is capable of inhibiting tPA. In order to gain a better insight into this specificity, docking studies were undertaken using the model of ETI, after manual correction of the scissile peptide bond towards the canonical conformation (Lamba *et al.*, 1996). This kind of docking studies will benefit significantly from a high-resolution structure of a Kunitz-type trypsin inhibitor in its complex with a cognate proteinase. Our highly refined model of STI:PPT complex provides an excellent template to build a homology model of ETI:tPA complex. The hypothetical models of both ETI:tPA and STI:tPA complexes were subject to energy minimization, after the following changes were made to STI to mimic ETI: removal of Asp1 and replacement of Phe2 by Val and Arg65 by Ala. The results of our modeling studies largely confirm those of the previous docking studies (Lamba *et al.*, 1996), which indicated that the exposed reactive site loop of ETI should slot into tPA's active-site cleft from the P3 to the P2′ residue. Furthermore, potentially significant ionic and hydrogen bonding interactions involving the residues outside the P3–P3′ segment have also been indicated by our studies (Figure 11(b)), in addition to the interactions between the reactive site loop of ETI from the P3 to P3′ residues and tPA described previously (Lamba *et al.*, 1996).

N-terminal Val1 ammonium group of ETI was suggested to be located close to the distal polar groups of Gln60′ and Glu60′A with the Val1 side-chain approaching the Arg39′ side-chain of tPA (Lamba *et al.*, 1996). Our modeling studies indicate that the Val1 side-chain is located above the ionically interacting pair of side-chains of Arg39′ and Glu60′A, but the side-chain of Gln60′ points away from the Val1 side-chain (Figure 11(b)). This interaction with Glu60′A explains the much weaker binding of the mutant ETI with an extra aspartic acid at the N terminus (Teixera *et al.*, 1994). This interaction is also partly responsible for STI and WCI being unable to inhibit tPA. STI has an extra aspartic acid residue at the N terminus and WCI two extra aspartic acid residues (Table 5). These extra residues of STI and WCI at the N terminus protrude from the body of the molecule, when compared with the ETI model. The protruding N termini of both STI and WCI may restrict their access to the active site of tPA by making a close contact with the characteristic insertion loop

Table 5. Residues of ETI involved in its interaction with tPA

| | | | P3 | P2 | P1 | P1′ | P2′ | P3′ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| (ETI No.) | | | 1 | 61 | 62 | 63 | 64 | 65 | 66 | 69 | 70 | 71 | 72 | 116 | 135 |
| ETI | | | V | R | L | R | S | A | F | D | D | D | K | K | K |
| ETIa | | | V | R | L | R | S | A | F | D | D | D | K | K | K |
| ETIb | | | E | P | F | R | S | Y | F | K | G | S | L | K | E |
| (STI No.) | −1 | 1 | 2 | 61 | 62 | 63 | 64 | 65 | 66 | 69 | 70 | 71 | 72 | 119 | 139 |
| STI | | D | F | P | Y | R | I | R | F | E | G | H | P | R | Q |
| WCI | D | D | D | Q | F | L | S | L | F | R | G | S | L | K | D |

The underlined residues are predicted to interact with tPA. The residues in boldface are critical for specific binding of tPA by ETI (and ETIa). Other residues are also listed to facilitate the discussion.

(60Loop) as well as by making a repulsive ionic interaction. ETIa from *Erythrina variegata*, belonging to group c, has the same valine residue at the N terminus as ETI and shows the inhibitory activity against tPA. In contrast, the N-terminal residue of ETIb is a glutamic acid and it does not inhibit tPA, even without extra negatively charged residues at the N terminus (Table 5). This is readily explained by the ionic interaction of the N-terminal Val1 of ETI with Glu60′A in tPA.

There is a pronounced difference in the surface features of STI and ETI, despite similarity in their overall three-dimensional structures (Figure 13). Another region of tPA important for inhibitor binding is 97Loop. This β-hairpin loop protrudes considerably from the active site of tPA. Together with Glu93′, the side-chains of Asp95′, Asp96′ and Asp97′ form a highly acidic surface patch in the tPA structure (Figure 13). The three residues Asp95′, Asp96′ and Asp97′ interact with the positively-charged side-chains of Lys72, Lys116 and Lys135 of ETI (Figure 13). In STI, WCI and ETIb, which do not show inhibitory activity toward tPA, two of these three lysines (Lys72 and Lys135) are replaced by neutral or negatively charged residues, while a positive charge is retained at the position corresponding to Lys116 of ETI (Table 5). The side-chain nitrogen atom of Lys72 makes salt bridges with side-chain oxygen atoms of Asp96′ and Asp97′ in 97Loop of tPA. The side-chain nitrogen atom of Lys116 in ETI also interacts with the side-chain of Asp96′ in tPA. However, this interaction may not be important in the selective inhibition of tPA by ETI, since the corresponding residue

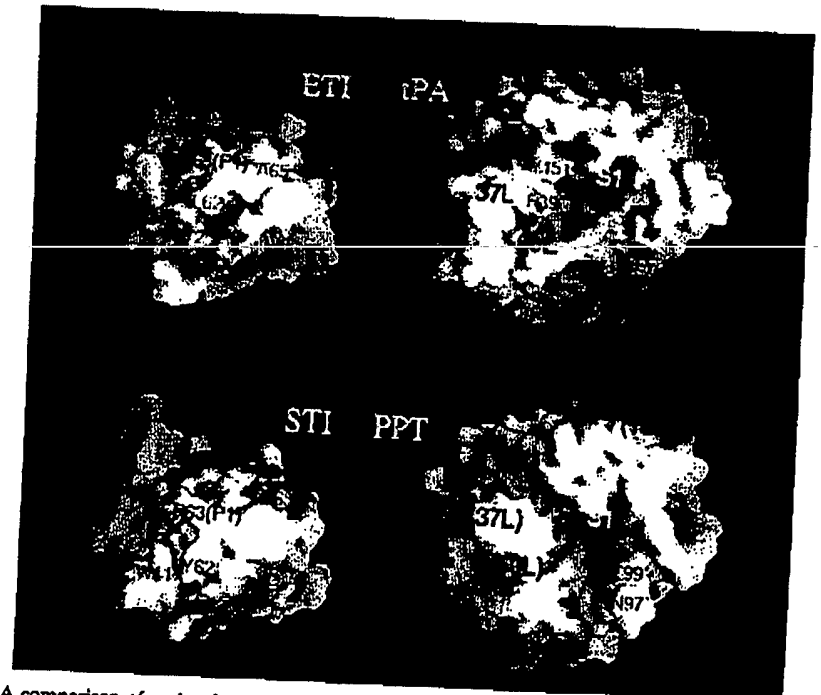


Figure 13. A comparison of molecular surfaces. Positively charged regions are blue and negatively charged regions red. The P1 residue, S1 pocket and some basic and acidic patches are labeled. This figure was generated using GRASP (Nicholls, 1992).

Arg119 in STI is predicted to make a similar interaction with Asp96′ of tPA. The side-chain of Lys135 in ETI makes at least three hydrogen bonds with the side-chains of Asp95′ and Thr98′ in addition to the ionic interaction with Asp95′ (Figure 11(b)).

A mutagenesis study on ETIa from *E. variegata* showed that the tPA inhibitory activity of the single mutants R61P (P3) and L62F (P2) is significantly reduced and, furthermore, the double mutant R61P/L62F (P3/P2) lacks the tPA inhibitory activity, despite retaining the trypsin inhibitory activity (Kouzuma *et al.*, 1997). These mutants were designed on the basis of the sequence difference between ETIa and ETIb (Table 5). The guanidyl group of Arg61 in ETI makes hydrogen bonds with the main-chain oxygen atoms of Arg174′ and Thr175′ in tPA (Figure 11(b)). The P3 residues of STI and WCI are proline and glutamine, respectively. Both crystal structures of STI:PPT(ortho) and STI:PPT(tetra) show that the side-chain of Tyr62 (P2) in STI makes a van der Waals contact with Leu99′ in PPT. Interestingly, the interacting side-chains are switched in the ETI:tPA complex and a comparable interaction is possible between the side-chains of Leu62 (P2) in ETI and Tyr99′ in tPA. A unique feature of ETI (and ETIa also), as compared with the inhibitors which do not inhibit tPA, is the presence of three consecutive acidic residues, Asp69, Asp70 and Asp71 (Table 5). However, its significance is not clear, since only Asp71 makes hydrogen bonds with the side-chain of Tyr99′. A mutant A65Y (P2′) of ETIa from *E. variegata* exhibited the same level of tPA inhibitory activity as the wild-type ETIa. (Kouzuma *et al.*, 1997). Modeling indicates that an aromatic ring of tyrosine at P2′ position can interact favorably with that of Tyr151′ in tPA through an edge–face interaction, the most common mode of aromatic–aromatic interaction (Burley & Petsko, 1985). Other residues of ETI which are possibly involved in the interaction with tPA are Gln11, Asn12, Gly13 and Phe111. Table 5 lists the residues of ETI which play an important role in its interaction with tPA.

## Materials and Methods

### Purification and crystallization

Kunitz-type soybean trypsin inhibitor (isoform Ti*, T-9003) and porcine pancreatic trypsin (T-7418) were purchased from Sigma. For the complex formation, STI (138 mg) and PPT (102 mg) were dissolved in 100 mM sodium chloride solution buffered at pH 8.0 with 100 mM Tris-HCl and were left for 2 h at 4°C. This solution was loaded on a Sephacryl S-200 HR gel filtration column (2.2 cm × 76.0 cm), which had been previously equilibrated with the same buffer. The fractions containing the STI:PPT complex were collected and concentrated to ~110 mg/ml concentration (1.6 ml volume) by ultrafiltration using a YM 10 membrane (Amicon) for crystallization. The protein concentration was estimated by measuring the absorbance at 280 nm, assuming a correspondence of 1.0 mg/ml concentration to the unit absorbance at 280 nm for the 1.0 cm path length. Crystallization was achieved by the hanging drop vapor diffusion method using 24-well tissue culture plates (Flow Laboratories) at room temperature (22 ± 1°C). The hanging drop was prepared by mixing equal volumes of the protein solution and the reservoir solution. The STI:PPT complex have been crystallized in two crystal forms. For the orthogonal form, the reservoir solution is 30% (v/v) PEG 400 and 100 mM lithium sulfate in 100 mM Hepes at a final pH of 7.67. Crystals of typical dimensions of 0.5 mm × 0.6 mm × 1.5 mm grew within a week. For the tetragonal form, the reservoir solution is 1.70 M ammonium sulfate and 10% (v/v) dioxane in 100 mM Mes at a final pH of 5.68. Crystals of typical dimensions of 2.0 mm × 2.0 mm × 0.4 mm grew within a week. The detailed crystallization condition for the free STI has been described previously (Lee *et al.*, 1993). Crystals of dimensions of 0.5 mm × 0.4 mm × 0.2 mm grew within about 25 days.

### X-ray diffraction studies

For all X-ray data collection, a crystal was mounted in a thin-walled glass capillary and both ends of the capillary were filled with the mother liquor and then sealed with wax. Three sets of data were collected at three places using different types of detectors. X-ray diffraction studies on the free STI have been described previously (Lee *et al.*, 1993). A summary of data collection is given in Table 6.

Data from an orthorhombic crystal of STI:PPT complex were collected on a MAR image plate X-ray detector at the beamline X12C of the National Synchrotron Light Source, Brookhaven National Laboratory (Sweet, 1989). All data were reduced and scaled using DENZO (Gewirth, 1993) and SCALEPACK (Otwinowski, 1992). The crystal diffracted to beyond 1.7 Å resolution using synchrotron X-rays. Despite much difference in crystallization conditions, our orthorhombic crystal of the complex is similar to the crystal previously grown by ethanol (Sweet *et al.*, 1974).

Data from a tetragonal crystal of STI:PPT complex were collected at 14°C using a Weissenberg camera for macromolecular crystallography at the BL-6A2 experimental station of Photon Factory, Japan (Sakabe, 1991). A Fuji image plate (20 × 40 cm) was placed at a distance of 429.7 mm from the crystal. The oscillation range per image plate was 3.0°, with a speed of 2.0°/sec and a coupling constant of 1.0°/min. An overlap of 0.5° was allowed between two contiguous image plates. Image plates were digitized by a Fuji BA100 scanner. Raw data were processed using the program WEIS (Higashi, 1989). The crystal belongs to the tetragonal space group $P4_12_12$ or $P4_32_12$. The former space group was later found to be correct on the basis of translation function calculations.

### Molecular replacement

All three structures were solved by the molecular replacement method (Crowther & Blow, 1967) using the programs X-PLOR (Brünger, 1992) and AMoRe (Navaza, 1994). Reflections with $F_o > 2\sigma_F$ were used throughout the molecular replacement and subsequent refinement calculations. First, a molecular replacement solution of the free STI structure was attempted, using the previously reported model of ETI as the probe structure (Onesti *et al.*, 1991; PDB ID code, 1TIE). A consistent

Table 6. Data collection statistics

| | STI:PPT(ortho) | STI:PPT(tetra) | STI |
|---|---|---|---|
| X-rays | 1.000 Å (NSLS[a]) | 1.000 Å (PF[b]) | CuKα (SNU[c]) |
| No. of crystals | 1 | 1 | 1 |
| Resolution limit (Å) | 1.75 | 1.80 | 2.1 |
| Space group | $P2_12_12_1$ | $P4_32_12$ | $P2_12_12_1$ |
| Cell axes (*a*, *b*, *c* in Å) | 58.91, 62.33, 151.46 | 62.45, 62.45, 229.11 | 39.42, 58.08, 96.35 |
| Asymmetric unit | 1 (STI:PPT) | 1 (STI:PPT) | 1 (STI) |
| $V_M$ (Å$^3$/Da) | 3.16 | 2.56 | 2.75 |
| Solvent content (%) | 61 | 52 | 55 |
| No. of total reflections | 419,433 | 118,553 | 27,868 |
| No. of unique reflections | 52,687 | 26,956 | 10,238 |
| $R_{merge}$ (%) | 7.8 | 7.9 | 4.9 |
| Data completeness (%) | 91.8 (50.0–1.75 Å)<br>68.6 (1.80–1.75 Å) | 62.0 (60.0–1.80 Å)<br>37.2 (1.90–1.80 Å) | 75.9 (33.0–2.10 Å)<br>26.7 (2.30–2.10 Å) |

[a] National Synchrotron Light Source, Brookhaven National Laboratory, USA.
[b] Photon Factory, Tsukuba, Japan
[c] Seoul National University, Seoul, Korea.

rotation and translation solution was obtained for different resolution ranges. And the refinement of the free STI model converged slowly to an *R*-factor of about 25% for 8.0 to 2.3 Å data. At this time, X-ray data from an orthorhombic crystal of STI:PPT complex became available and its structure was solved by molecular replacement using the model of porcine β-trypsin refined at 1.6 Å (Huang *et al.*, 1993; PDB ID code, 1MCT). After orienting and positioning the PPT model alone according to the molecular replacement solution, an *R*-factor of 42.4% and a correlation coefficient of 0.50 were obtained for 10.0 to 4.0 Å data. A model for the complex was then generated by positioning the partially refined model of free STI through a careful inspection of the electron density and also the interaction between trypsin and the inhibitor. Electron density for the STI part in the complex improved during the successive refinement steps and the complex structure was refined to an *R*-factor of 18.9% for 8.0 to 1.75 Å data. Now, this highly refined model of STI in the complex was used in turn for a structure re-determination of free STI. It gave a clear rotation solution ($\alpha = 88.10°$, $\beta = 31.45°$, $\gamma = 10.49°$) and a translation solution ($x = 0.2719$, $y = 0.2867$, $z = 0.3261$ in fractional coordinates). After the STI model was oriented and positioned according to the above solution, it gave an *R*-factor of 32.6% and a correlation coefficient of 0.73 for 12.0 to 4.0 Å data.

The starting model for STI:PPT(tetra) was the refined model of STI:PPT(ortho). The search for the molecular replacement solution was performed similarly. The correct space group was determined to be $P4_12_12$ by translation function calculations. For 10.0 to 4.0 Å data, an *R*-factor of 35.6% and a correlation coefficient of 0.51 were obtained for $P4_12_12$, compared with an *R*-factor of 46.0% and a correlation coefficient of 0.26 for $P4_32_12$.

## Refinement

The STI:PPT(ortho) model was subject to refinement by the program X-PLOR version 3.1 (Brünger, 1992). Initially, a rigid-body refinement was carried out with 8.0 to 6.0 Å data to further improve the positional and orientational parameters. The high-resolution limit of the diffraction data was increased stepwise from 6.0 to 3.5 Å. The *R*-factor at this stage was 36.9 % for 8.0 to 3.5 Å data. Next, atomic positions were refined by the conventional conjugate gradient minimization, with higher resolution data being added in steps. During this stage, the *R*-factor dropped to 27.5% for 8.0 to 2.5 Å data. This model was then subject to a simulated annealing refinement, employing the standard slow cooling protocol: from 3000 K to 300 K (time-step 0.5 fs; decrement of temperature 25 K; number of steps at each temperature 50). Solvent molecules, modeled as water, were located as high peaks in the $(F_o - F_c)$ maps and were included in the subsequent rounds of refinement. A water molecule was removed when its temperature factor (*B*-factor) exceeded 60 Å$^2$. Higher resolution reflections up to 1.75 Å were added in steps. Further simulated annealing refinement and the refinement of isotropic *B*-factors for individual atoms gave the final *R*-factor of 18.9% for 31,038 unique reflections with $F_o > 2\sigma_F$ in the range 8.0 to 1.75 Å. The final free *R*-factor is 21.4%. The refinements of STI:PPT(tetra) and free STI models were performed similarly. A summary of the refinement statistics is given in Table 1.

## Model building and structure analysis

The program CHAIN version 5.4 running on a Silicon Graphics Indigo$^2$ XZ workstation was used for model rebuilding (Sack, 1988) and the model stereochemistry was assessed by PROCHECK (Laskowski *et al.*, 1993). Structural comparisons were made using the program LSQKAB in the CCP4 program package (SERC Daresbury Laboratory, 1994). The orthorhombic crystal structure of STI:PPT complex was used as a template to build a homology models of tPA complexes with inhibitors. In homology modeling of ETI:tPA complex, the structure of recombinant two-chain human tPA (Lamba *et al.*, 1996; PDB ID code, 1RTF) was superposed with PPT (with an r.m.s. deviation of 0.92 Å for 181 $C^\alpha$ atom pairs). The ETI structure (Onesti *et al.*, 1991; PDB ID code, 1TIE) was also superposed with STI. The resulting models of the complexes between the inhibitors and tPA were then subject to energy minimization by the CHARMM dynamics in the program package QUANTA (Molecular Simulations, Inc.). Programs CHAIN version 5.4 (Sack, 1988), GRASP (Nicholls, 1992), and MOLSCRIPT version 2.1.4 (Kraulis, 1991) were used on a Silicon Graphics Indigo$^2$ XZ workstation for model display. The atomic coordinates have been deposited with the Protein Data Bank (ID codes: 1AVU for the free STI, 1AVW for the STI:PPT(ortho), and 1AVX for the STI:PPT(tetra)).

## Acknowledgments

This work was supported by the Center for Molecular Catalysis, Seoul National University, S.N.U. Dae Woo Research Fund (95-05-2057), Korea Science and Engineering Foundation (95-0501-07-01-3), and the Korea Ministry of Education, Basic Sciences Research Institute. We thank Dr Robert M. Sweet of the beamline X12C at National Synchrotron Light Source, Brookhaven National Laboratory, USA. and Professor N. Sakabe and Dr M. Suzuki of the beamline BL-6A2, Photon Factory, KEK, Japan for their assistance with data collection. The Inter-University Center for Natural Science Research Facilities, Seoul National University is also thanked for providing the X-ray equipment, which is partially supported by the Specialization Fund from KOSEF.

## References

Baillargeon, M. W., Laskowski, M., Neves, D. E., Porubcan, M. A., Santini, R. E. & Markley, J. L. (1980). Soybean trypsin inhibitor (Kunitz) and its complex with trypsin. Carbon-13 nuclear magnetic resonance studies of the reactive site arginine. *Biochemistry*, 19, 5703–5710.

Blow, D. M., Janin, J. & Sweet, R. M. (1974). Mode of action of soybean trypsin inhibitor (Kunitz) as a model for specific protein–protein interactions. *Nature*, 249, 54–57.

Bode, W. & Huber, R. (1992). Natural protein proteinase inhibitors and their interactions with proteinases. *Eur. J. Biochem.* 204, 433–451.

Brünger, A. T. (1992). *X-PLOR Manual, Version 3.1: A System for Crystallography and NMR*, Yale University Press, New Haven.

Burley, S. K. & Petsko, G. A. (1985). Aromatic–aromatic interaction: a mechanism of protein structure stabilization. *Science*, 229, 23–28.

Collen, D. & Lijnen, H. R. (1995). Molecular basis of fibrinolysis, as relevant for thrombolytic therapy. *Thromb. Haemostasis*, 74, 167–171.

Crowther, R. A. & Blow, D. M. (1967). A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallog.* 23, 544–548.

Dattagupta, J. K., Podder, A., Chakrabarti, C., Sen, U., Dutta, S. K. & Singh, M. (1996). Structure of a Kunitz-type chymotrypsin inhibitor from winged bean seeds at 2.95 Å resolution. *Acta Crystallog. sect. D*, 52, 521–528.

Eriksson, A. E., Causens, L. S., Weaver, L. H. & Matthews, B. W. (1991). Three-dimensional structure of human basic fibroblast growth factor. *Proc. Natl Acad. Sci. USA*, 88, 3441–3445.

Finzel, B. C., Clancy, L. L., Holland, D. R., Muchmore, S. W., Watenpaugh, K. D. & Einspahr, H. M. (1989). Crystal structure of recombinant human interleukin-1β at 2.0 Å resolution. *J. Mol. Biol.* 209, 779–791.

Gewirth, D. (1993). *The Denzo Manual: An Oscillation Data Processing Program for Macromolecular Crystallography by Zbyszek Otwinoski*, Yale University, New Haven, CT.

Graves, B. J., Hatada, M. H., Hendrickson, W. A., Miller, J. K., Madison, V. S. & Satow, Y. (1990). Structure of interleukin-1α at 2.7 Å resolution. *Biochemistry*, 29, 2679–2684.

Habazetti, J., Gondol, D., Witscheck, R., Ottewski, J., Schleicher, M. & Holak, T. A. (1992). Structure of hisactophilin is similar to interleukin-1β and fibroblast growth factor. *Nature*, 359, 855–858.

Heussen, C., Joubert, F. & Dowdle, E. B. (1984). Purification of human tissue plasminogen activator with *Erythrina* trypsin inhibitor. *J. Biol. Chem.* 259, 11635–11638.

Higashi, T. (1989). The processing of diffraction data taken on a screenless Weissenberg camera for macromolecular crystallography. *J. Appl. Crystallog.* 22, 9–18.

Huang, Q., Liu, S. & Tang, Y. (1993). Refined 1.6 Å resolution structure of the complex formed between porcine β-trypsin and MCTI-A, a trypsin inhibitor of the squash family. *J. Mol. Biol.* 229, 1022–1036.

Kim, S.-H., Hara, S., Hase, S., Ikenaka, T., Toda, H., Kitamura, K. & Kaizuma, N. (1985). Comparative study on amino acid sequences of Kunitz-type soybean trypsin inhibitors, Ti[a], Ti[b], and Ti[c]. *J. Biochem.* 98, 435–448.

Kortt, A. A., Strike, P. M. & De Jersey, J. (1989). Amino acid sequence of a crystalline seed albumin (winged bean albumin-1) from *Psophocarpus tetragonalobus* (L.) DC. *Eur. J. Biochem.* 181, 403–408.

Kouzuma, Y., Yamasaki, N. & Kimura, M. (1997). The tissue-type plasminogen activator inhibitor ETIa from *Erythrina variegata*: structural basis for the inhibitory activity by cloning, expression, and mutagenesis of the cDNA encoding ETIa. *J. Biochem.* 121, 456–463.

Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* 24, 946–950.

Kunitz, M. (1947a). Crystalline soybean trypsin inhibitor. *J. Gen. Physiol.* 30, 291–310.

Kunitz, M. (1947b). Isolation of a crystalline protein compound of trypsin and soybean trypsin inhibitor. *J. Gen. Physiol.* 30, 311–320.

Lamba, D., Bauer, M., Huber, R., Fischer, S., Rudolph, R., Kohnert, U. & Bode, W. (1996). The 2.3 Å crystal structure of the catalytic domain of recombinant two-chain human tissue-type plasminogen activator. *J. Mol. Biol.* 258, 117–135.

Laskowski, M., Jr., Kato, I., Leary, T. R., Schrode, J. & Sealock, R. W. (1974). Evolution of specific protein proteinase inhibitors. In *Proteinase Inhibitors: Bayer Symposium V* (Fritz, H., Tschesche, H., Greene, L. J. & Truscheit, E., eds), pp. 597–611, Springer, Berlin.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* 26, 283–291.

Leah, R. & Mundy, J. (1989). The bifunctional α-amylase/subtilisin inhibitor of barley: nucleotide sequence and patterns of seed-specific expression. *Plant Mol. Biol.* 12, 673–682.

Lee, J., Song, H. K., Hwang, K. Y., Kim, K. K. & Suh, S. W. (1993). Crystallization and preliminary X-ray crystallographic study of Kunitz-type soybean trypsin inhibitor. *Mol. Cells*, 3, 335–337.

Levitt, M. & Perutz, M. F. (1988). Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* 201, 751–754.

Loebermann, H., Tokuoka, R., Deisenhofer, J. & Huber, R. (1984). Human $\alpha_1$-proteinase inhibitor. *J. Mol. Biol.* 177, 531–556.

Mareš, M., Meloun, B., Pavlik, M., Kostka, V. & Baudyš, M. (1989). Primary structure of cathepsin D inhibitor from potatoes and its structure relationship to soybean trypsin inhibitor family. *FEBS Letters*, 251, 94–98.

Marquart, M., Walter, J., Deisenhofer, J., Bode, W. & Huber, R. (1983). The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallog. sect. B*, 39, 480–490.

McLachlan, A. D. (1979). 3-fold structural pattern in the soybean trypsin inhibitor (Kunitz). *J. Mol. Biol.* 133, 557–563.

Murzin, A. G., Lesk, A. M. & Chothia, C. (1992). β-trefoil fold. *J. Mol. Biol.* 223, 531–543.

Navaza, J. (1994). AMoRe: an automatic package for molecular replacement. *Acta Crystallog. sect. A*, 50, 157–163.

Nicholls, A. (1992). *GRASP: Graphical Representation and Analysis of Surface Properties*, Columbia University, New York.

Onesti, S., Brick, P. & Blow, D. M. (1991). Crystal structure of a Kunitz-type trypsin inhibitor from *Erythrina caffra* seeds. *J. Mol. Biol.* 217, 153–176.

Otwinoski, Z. (1992). *The Scalepack Manual*, Yale University, New Haven, CT.

Perona, J. J., Tsu, C. A., Craik, C. S. & Fletterick, R. J. (1993). Crystal structures of rat anionic trypsin complexed with the protein inhibitors APPI and BPTI. *J. Mol. Biol.* 230, 919–933.

Priestle, J. P., Schär, H.-P. & Grütter, M. (1988). Crystal structure of the cytokine interleukin-1β. *EMBO J.* 7, 339–343.

Read, R. J. & James, M. N. G. (1986). Introduction to protein inhibitors: X-ray crystallography. In *Proteinase Inhibitors* (Barrett, A. J. & Salvesen, G., eds), pp. 301–335, Elsevier, Amsterdam.

Read, R. J., Fujinaga, M., Sielecki, A. R. & James, M. N. G. (1983). Structure of the complex of *Streptomyces griseus* protease B and the third domain of the turkey ovomucoid inhibitor at 1.8 Å resolution. *Biochemistry*, 22, 4420–4433.

Ringe, D., Seaton, B. A., Gelb, M. H. & Abeles, R. H. (1985). Inactivation of chymotrypsin by 5-benzyl-6-chloro-2-pyrone: $^{13}C$ NMR and X-ray diffraction analyses of the inactivator-enzyme complex. *Biochemistry*, 24, 64–68.

Rutenber, E. & Robertus, J. D. (1991). Structure of ricin B-chain at 2.5 Å resolution. *Proteins: Struct. Funct. Genet.* 10, 260–269.

Sack, J. S. (1988). CHAIN: a crystallographic modeling program. *J. Mol. Graphics*, 6, 244–245.

Sakabe, N. (1991). X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchrotron radiation. *Nucl. Instrum. Methods A*, 303, 448–463.

SERC Daresbury Laboratory (1994). The CCP4 suit: Programs for protein crystallography. *Acta Crystallog. sect. D*, 50, 760–763.

Song, S. I., Kim, C. H., Baek, S. J. & Choi, Y. D. (1991). Molecular cloning and nucleotide sequencing of cDNA encoding the precursor for soybean trypsin inhibitor (Kunitz type). *Mol. Cells*, 1, 317–324.

Song, H. K., Lee, K. N., Kwon, K.-S., Yu, M.-H. & Suh, S. W. (1995). Crystal structure of an uncleaved $\alpha_1$-antitrypsin reveals the conformation of its inhibitory reactive loop. *FEBS Letters*, 377, 150–154.

Sweet, R. M. (1989). Facilities at the National Synchrotron Light Source at Brookhaven National Laboratory. In *Synchrotron Radiation in Structural Biology* (Sweet, R. M. & Woodhead, A. D., eds), pp. 63–66, Plenum Press, New York.

Sweet, R. M., Wright, H. T., Janin, J., Chothia, C. H. & Blow, D. M. (1974). Crystal structure of the complex of porcine trypsin with soybean trypsin inhibitor (Kunitz) at 2.6 Å resolution. *Biochemistry*, 13, 1599–1608.

Takeuchi, Y., Satow, Y., Nakamura, K. T. & Mitsui, Y. (1991). Refined crystal structure of the complex of subtilisin BPN' and Streptomyces subtilisin inhibitor at 1.8 Å resolution. *J. Mol. Biol.* 221, 309–325.

Teixeira, A. V., Dowdle, E. B. D. & Botes, D. P. (1994). Site-directed mutagenesis of the synthetic *Erythrina* trypsin/tissue plasminogen activator (t-PA) inhibitor encoding-gene to compare the interaction of *Erythrina caffra* and soybean trypsin inhibitor with tPA. *Biochim. Biophys. Acta*, 1217, 23–28.

Wlodawer, A., Deisenhofer, J. & Huber, R. (1987). Comaprison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol.* 193, 145–156.

Yamamoto, M., Hara, S. & Ikenaka, T. (1983). Amino acid sequence of two trypsin inhibitors from winged bean seeds (*Psophocarpus tetragonalobus* (L.) DC). *J. Biochem.* 94, 849–863.

Zemke, K. J., Müller-Fahrnow, A., Jany, K.-D., Pal, G. P. & Saenger, W. (1991). The three-dimensional structure of the bifunctional proteinase K/α-amylase inhibitor from wheat (PKI3) at 2.5 Å resolution. *FEBS Letters*, 279, 240–247.

Zhang, J., Cousens, L. S., Barr, P. J. & Sprang, S. R. (1991). Three-dimensional structure of human basic fibroblast growth factor, a structural homolog of interleukin-1β. *Proc. Natl Acad. Sci. USA*, 88, 3446–3450.

Zhu, M., Komiya, H., Chirino, A., Faham, S., Fox, G. M., Arakawa, T., Hsu, B. T. & Rees, D. C. (1991). Three-dimensional structures of acidic and basic fibroblast growth factors. *Science*, 251, 90–93.

*Edited by R. Huber*

*(Received 30 June 1997; received in revised form 7 October 1997; accepted 9 October 1997)*

J. Biochem. 115, 397-404 (1994)

# Amino Acid Sequences of Kunitz Family Subtilisin Inhibitors from Seeds of *Canavalia lineata*

**Shigeyuki Terada, Hideki Katayama, Koichiro Noda, Satoshi Fujimura, and Eiji Kimoto**

*Department of Chemistry, Faculty of Science, Fukuoka University, 8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-01*

Received for publication, September 14, 1993

**The amino acid sequences of two subtilisin inhibitors (CLSI-II and -III) from *Canavalia lineata* seeds were determined by manual Edman degradation using the DABITC/PITC double coupling method after enzymatic digestions and CNBr degradation. CLSI-II and -III consist of 190 and 183 amino acids, respectively, and have identical amino acid sequences except for in the C-terminal regions: an elongated sequence, Gly-Thr-Ile-Arg-Ser-Asp-Gly, was found at the C-terminus of CLSI-II. A short-chain analog protein without an N-terminal Asn residue was also detected in each inhibitor preparation. The inhibitors showed significant homology to Kunitz type inhibitors, but differed from them with respect to the half-cystine content. Among five half-cystine residues present in CLSI-III, two disulfide bonds link $Cys^{44}$ to $Cys^{88}$ and $Cys^{142}$ to $Cys^{149}$, $Cys^{106}$ being present as a free cysteine residue. Phenylglyoxal treatment abolished the inhibitory activity of CLSI-III, indicating the participation of an Arg residue in the interaction with the enzyme. The reactive-site peptide bond was deduced to be $Arg^{68}$-$Gly^{69}$.**

**Key words: amino acid sequences, *Canavalia lineata*, Kunitz inhibitor family, subtilisin inhibitors.**

Subtilisin inhibitory activities have been detected in many legume seeds (*1-3*). Subtilisin inhibitors from broad bean (*4*) and adzuki bean (*5*) have already been sequenced. These have low $M_r$ of about 8,000 and belong to the potato I type inhibitor family. Bifunctional inhibitors that inhibit both subtilisin and $\alpha$-amylase simultaneously have also been purified from cereal plants, and their sequences showed homology with Kunitz type inhibitors (*6-8*). Kunitz type proteins are characterized by their molecular size ($M_r$ 21,000-22,000), the presence of two disulfide bonds (4 half-cystine residues), and the single reactive site for serine proteases (*9*).

Subtilisin isoinhibitors of $M_r$ 8,000 have been purified from jack bean (*Canavalia ensiformis*) (*3*), but their amino acid sequences have not been reported yet. Recently, we isolated and characterized two subtilisin inhibitors, CLSI-II and -III, from seeds of a similar bean, *C. lineata* (*10*). These were considered to be Kunitz type inhibitors in spite of the presence of 5 half-cystine residues, unlike in all other Kunitz inhibitors. We have herein confirmed the classification of CLSI-II and -III by determining their complete amino acid sequences. These are the first Kunitz type subtilisin inhibitors from legume seeds of which the primary structures have been elucidated.

## MATERIALS AND METHODS

This portion of the paper, in addition to Figs. 1S-10S and Tables IS-VIS, is presented as a miniprint at the end of this paper.

Abbreviations: CLSI, *Canavalia lineata* subtilisin inhibitor; DABITC/PITC, 4-*N*,*N*-dimethylaminoazobenzene-4'-isothiocyanate/phenylisothiocyanate; STI, Kunitz soybean trypsin inhibitor; PE, *S*-pyridylethylated; CM, *S*-carboxymethylated; TPCK, 1-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone.

## RESULTS

*N-Terminal Amino Acid Sequences of PE-CLSI-II and -III*—The N-terminal residues of PE-CLSI-II and PE-CLSI-III were analyzed. In each cycle of manual Edman degradation, almost equal amounts of two amino acids were detected for both inhibitors. For example, PE-CLSI-II gave the sequence of (N,D)-(D,V)-(V,D)-(D,V)-(V)-(V,M)-(M,D)-(D,A)……. This clearly indicates that two peptide chains are present in the inhibitor preparation, one of which is devoid of the N-terminal Asn residue. The same phenomenon was observed on the sequencing of all peptides containing the N-terminal parts of the inhibitors. Nevertheless, the N-terminal 24 residues of CLSI-II and 18 residues of CLSI-III were identified (Figs. 1 and 2), and the sequences were identical.

*Enzymatic and Chemical Digestion of the Reduced and Alkylated Inhibitors*—Peptides generated on lysyl endopeptidase digestion of PE-CLSI-II were fractionated by reverse-phase HPLC on a C-18 column (Fig. 1S) and then subjected to amino acid analysis (Tables IS). The peptide peaks named K5/6 and K7 represented mixtures of two peptides. K5/6 was rechromatographed on the same column using different gradient conditions to give two peptides, K5 and K6. On the other hand, K7 was further digested with TPCK-trypsin without separation of the constituent peptides. The elution profile of the tryptic digest is shown in Fig. 2S. The amino acid compositions of the purified peptides (K7T1-K7T5) are also cited in Table IS. The digestion with V8 protease gave ten major peaks

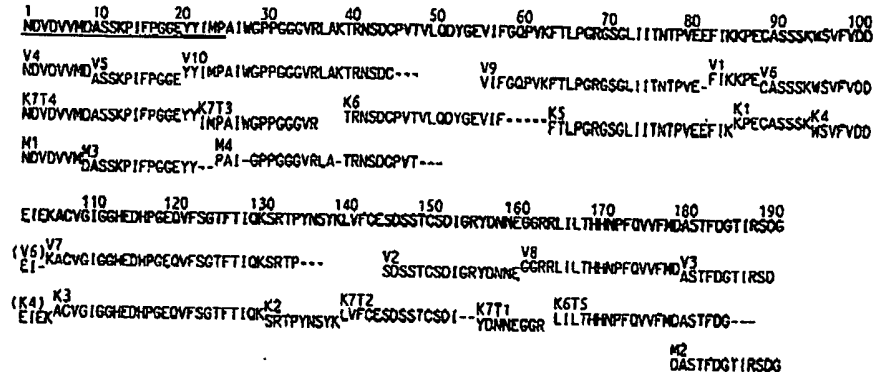

Fig. 1. Summary of amino acid sequence analysis of CLSI-II. Peptides derived from CLSI-II on digestions with trypsin (T), lysyl endopeptidase (K), *S. aureus* V8 protease (V), and CNBr (M) are shown above the sequences. Residues not identified are indicated by dashes. The N-terminal part of the inhibitor was determined by directly sequencing PE-CLSI-II and is shown by an underline.

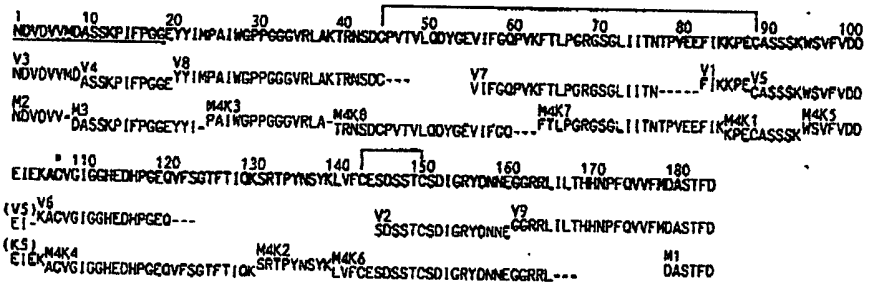

Fig. 2. Summary of amino acid sequence analysis of CLSI-III. Residues not identified are indicated by dashes. The peptide nomenclature is the same as in Fig. 1. Two disulfide bridges and a single cysteine residue are shown by solid lines and an asterisk, respectively.

(V1-V10) on HPLC (Fig. 3S). The amino acid compositions of the peptides are summarized in Table IIS. Chemical cleavage with CNBr gave four peptides, namely, M1, M2, M3, and M4, on fractionation by reverse-phase HPLC (Fig. 4S). Table IIIS shows the amino acid compositions of these peptides.

A slightly modified procedure was employed to digest CLSI-III. PE-CLSI-III was subjected to CNBr cleavage and the peptides were purified as described above (Fig. 5S). A large peptide, M4 in Fig. 5S, possessing the sequence of PAIWGPPVRLAKTRNSDCPV----, was further digested with lysyl endopeptidase to give eight peptides, M4K1-M4K8 (Fig. 6S). CM-CLSI-III was used instead of the PE-inhibitor for V8 protease digestion (Fig. 7S). The amino acid compositions of peptides derived from CLSI-III are cited in Tables IVS and VS.

*Complete Amino Acid Sequences of CLSI-II and -III*—The complete amino acid sequences of CLSI-II and -III are shown in Figs. 1 and 2, together with details of the overlapping peptides from which they were deduced. The locations of peptide K2 in CLSI-II and M4K2 in CLSI-III were determined to be exclusively at positions ranging from 130 to 138. The sequence of CLSI-II contained 190 amino acid residues corresponding to 20,770 Da, while CLSI-III consisted of 183 residues corresponding to 20,023 Da. The total numbers of amino acid residues in the sequences were in good agreement with the amino acid compositions of the two inhibitors, as described in a preceding paper (*10*). The sequences of the two inhibitors are different from each other only in the C-terminal parts of their molecules. The heptapeptide, Gly[184]-Thr-Ile-Arg-Ser-Asp-Gly[190], is attached at the C-terminal of CLSI-II.

*Location of the Disulfide Bridges*—Five cysteine residues are located at positions 44, 88, 142, 149, and 106 in

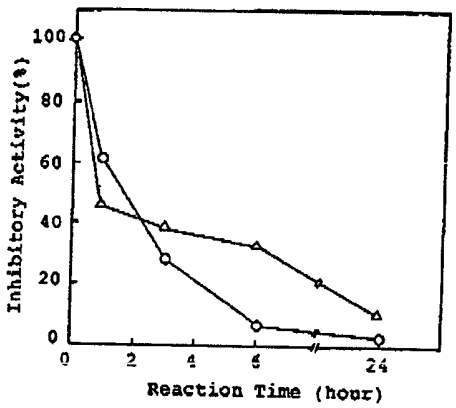


Fig. 3. Modification of arginine residues of CLSI-III and STI. The inhibitors were treated with 0.15% phenylglyoxal in 0.2 M *N*-ethylmorpholine-acetic acid buffer (pH 8.0) at 25°C for various periods, and then the remaining inhibitory activities of CLSI-III (○) and STI (△) toward subtilisin BPN′ and trypsin, respectively, were determined.

both inhibitors. To locate the disulfide bridges, unreduced CLSI-III was digested directly with trypsin. On reverse-phase HPLC, three cystine-containing peptides, namely, NT1, NT2, and NT3, were obtained (Fig. 8S). The amino acid compositions of NT1, NT2, and NT3 corresponded to residues 139-154, 41-62 and 85-93, and 105-129, respectively (Table VIS). The sequences of these peptides were also confirmed by Edman degradation to be as follows:

NT1, $^{139}$L-V-F-X-E-S-D-S-S-T-X-S-D-I-G-R$^{154}$

NT2, $^{41}$N-S-D-X-P-V-T-V-L-Q-D-$^{51}$
K-P-E-X-A-S-S-S-K
$_{85}$ $_{98}$

NT3, $^{105}$A-X-V-G-I-G-G-H-$^{112}$

The results indicate that two disulfide bonds link Cys$^{44}$ to Cys$^{88}$ and Cys$^{142}$ to Cys$^{149}$, Cys$^{106}$ being present as a free cysteine.

*Modification of Arginine Residues*—The relation between the reaction time with phenylglyoxal and the residual activity of CLSI-III is shown in Fig. 3. The subtilisin inhibitory activity decreased significantly with the reaction time. The inhibitor incubated similarly without phenylglyoxal retained the full activity after 24 h. A soybean trypsin inhibitor (STI) possessing an arginine residue at the $P_1$ position of the reactive site (*19*) was treated similarly for comparison. The trypsin inhibitory activity decreased with the progress of modification of the arginine residue. This suggests that the arginine residue of CLSI-III might play an important role in the interaction with an enzyme.

*Modification of the Reactive Site*—To directly locate the reactive site, CLSI-III was treated with 68 mol% subtilisin BPN′ for 1 h at pH 7.6. The inhibitor was separated from the enzyme by reverse-phase HPLC on a Biofine RPC-SC18 column, and then reduced and *S*-pyridylethylated. Fractionation of PE-CLSI-III gave three peaks on the same column (Fig. 9S). The yield of the modified inhibitor was about 5%. The amino- and carboxy-terminal sequences of fraction C were determined to be Gly-Ser-Gly-Leu-Ile-Ile-Thr-Asn-Thr- and -Ala-Ser-Thr-Phe-Asp, respectively. Fractions N and I gave Asn and Asp as the amino-terminal amino acids. The amino acid analyses showed that fraction I is intact PE-CLSI-III, and that N and C correspond to residues 1–68 and 69–183, respectively (data not shown). This suggests that the reactive site of CLSI is Arg$^{68}$-Gly$^{69}$.

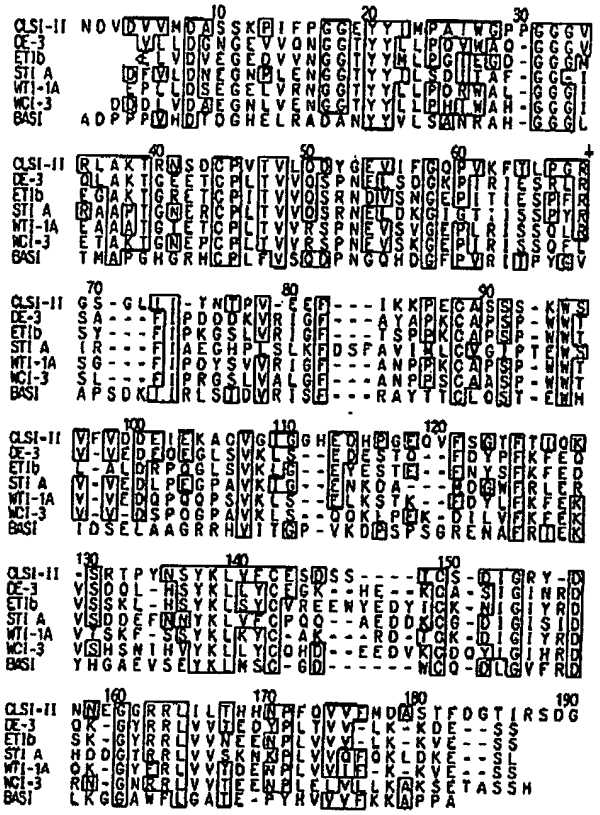


Fig. 4. Comparison of the amino acid sequences of Kunitz type inhibitors. DE-3, *Erythrina caffra* trypsin inhibitor (*24*); ETIb, *E. variegata* trypsin inhibitor (*25*); STI A, soybean trypsin inhibitor (*26*); WTI-1A, winged bean trypsin inhibitor (*27*); (*f*) WCI-3, winged bean chymotrypsin inhibitor (*28*); BASI, barley subtilisin/*α*-amylase inhibitor (*7*); ↓, reactive site. Residues identical to in CLSI-II are enclosed in solid boxes.

## DISCUSSION

We have determined the primary structures of two subtilisin inhibitors from seeds of *C. lineata*. These are the first Kunitz type subtilisin inhibitors from legume seeds of which the complete structures have been elucidated. The presence of a protein lacking the N-terminal asparagine residue was recognized for both inhibitors. Since *Canavalia* seeds contain a significant amount of a protease which specifically cleaves at the Asn-X peptide bond (*20*), such an enzyme might participate in the generation of the short-chain proteins.

CLSI-II and -III are isoinhibitors with different chain lengths. Isoinhibitors with an identical core sequence but different N-terminals have been found in many legume seeds (*21, 22*). The *C. lineata* subtilisin inhibitors in the present work as well as Bowman-Birk type trypsin inhibitors (*23*) from the same seed showed structural heterogeneity at their C-termini. The wider inhibitory spectrum and higher stability as to extreme pH of CLSI-II compared to CLSI-III should be due to the extended peptide chain at the C-terminus. However, it is not clear how the elongated residues endow the inhibitor with such properties.

CLSI-II and -III contain five Cys residues, four of which are linked by disulfide bridges. Cys$^{106}$ is assumed to be a cysteine residue and might be involved in the formation of dimeric molecules through an intermolecular disulfide bond when the inhibitors are stored for a long period in solution (*10*).

In Fig. 4, the amino acid sequences of nine known Kunitz type inhibitors are compared. CLSI-II showed significant homology with those of the Kunitz type inhibitors (*7, 24–28*). There are 21 invariant amino acid residues among these inhibitors. The relative locations of half-cystine residues are identical in all the inhibitors except for Cys$^{106}$ of CLSI. The *C. lineata* inhibitor (CLSI-II) exhibits the greatest homology (about 30% identity of residues) with *Erythrina caffra* DE-3 (*24*). Interestingly, the homology between CLSI and cereal subtilisin/*α*-amylase inhibitors (*6–8*) was relatively low (about 23%), despite their similar enzyme specificities. Limited but significant homology was also detected between Kunitz inhibitors and several storage proteins such as sporamins in sweet potato (*29*), miraculin in sweetberry (*30*), and seed albumin in winged bean (*31*).

We identified the reactive site of the inhibitor as Arg$^{68}$-Gly$^{69}$ by comparison with other inhibitors. The decrease in inhibitory activity upon modification of the Arg

residue(s) and the selective cleavage at this bond of CLSI-III by subtilisin support this conclusion. In spite of the occurrence of an Arg residue at the $P_1$ site, CLSI did not inhibit trypsin. Subtilisin has a hydrophobic pocket forming a subsite, $S_4$, and the $S_4$-$P_4$ interaction was shown to be important in the formation of an enzyme-substrate (inhibitor) complex (*32, 33*). The presence of a bulky hydrophobic residue, Leu, at the $P_4$ sites of CLSI-II and -III favors the interaction with the hydrophobic $S_4$ site of subtilisin. In contrast, trypsin has only subsite $S_1$-$S_3$ and no counterpart of the $S_4$ pocket of subtilisin (*33*). Thus, a bulky side chain at the $P_4$ site may interfere with the binding of inhibitors to the enzyme. In fact, all known Kunitz type trypsin inhibitors bear either Ser or Thr at the $P_4$ site.

Plots of hydropathy profiles (*34*) were compared among CLSI-III, *E. caffra* DE-3, and STI. The subtilisin inhibitor showed a quite similar pattern to the trypsin inhibitors except in the region (residues 50–90) including the reactive site (Fig. 10S). This might suggest the similarity of the secondary and tertiary structures of these Kunitz inhibitors, irrespective of the limited sequence homology. As described previously, a single cysteine residue ($Cys^{106}$) of CLSI was not involved in the interaction with the enzyme (*10*). If CLSI has a similar tertiary structure to other Kunitz inhibitors (*35*), it is deduced that this residue is located apart from the reactive site.

## REFERENCES


1. González, E., Callejas, A., Seidl, D.S., & Jaffé W.G. (1979) *J. Agric. Food Chem.* 27, 912-913
2. Chavan, J.K. & Hejgaard, J. (1981) *J. Sci. Food Agric.* 32, 857-862
3. Seidl, D.S., Lorenzo, P., Pinelli, B.E., & Tovar, J. (1988) *Biol. Chem. Hoppe-Seyler* 369, Suppl., 69-73
4. Svendsen, I., Hejgaard, J., & Chavan, J.K. (1984) *Carlsberg Res. Commun.* 49, 493-502
5. Nozawa, H., Yamagata, H., Aizono, Y., Yoshikawa, M., & Iwasaki, T. (1989) *J. Biochem.* 106, 1003-1008
6. Maeda, K. (1986) *Biochim. Biophys. Acta* 871, 250-256
7. Svendsen, I., Hejgaard, J., & Mundy, J. (1986) *Carlsberg Res. Commun.* 51, 43-50
8. Ohtsubo, K. & Richardson, M. (1992) *FEBS Lett.* 309, 68-72
9. Richardson, M. (1991) in *Methods in Plant Biochemistry* (Dey, P.M. & Harborne, J.B., eds.) Vol. 5, pp. 259-305, Academic Press, New York
10. Terada, S., Fujimura, S., Katayama, H., Nagasawa, M., & Kimoto, E. (1994) *J. Biochem.*, in press
11. Hermondson, M.A., Ericsson, L.H., Neurath, H., & Walsh, K.A. (1973) *Biochemistry* 12, 3146-3153
12. Hirs, C.H.W. (1967) in *Methods in Enzymology* (Hirs, C.H.W., ed.) Vol. 11, pp. 197-203, Academic Press, New York
13. Crestfield, A.M., Moore, S., & Stein, W.H. (1963) *J. Biol. Chem.* 238, 622-627
14. Matsubara, H. & Sasaki, R.M. (1969) *Biochem. Biophys. Res. Commun.* 35, 175-181
15. Chang, J.Y., Brauer, D., & Wittmann-Liebold, B. (1978) *FEBS Lett.* 93, 205-214
16. Chang, J.Y. (1981) *Biochem. J.* 199, 557-564
17. Takahashi, K. (1968) *J. Biol. Chem.* 243, 6171-6179
18. Ambler, R.P. (1972) in *Methods in Enzymology* (Hirs, C.H.W., ed.) Vol. 25, pp. 143-145, Academic Press, New York
19. Osawa, K. & Laskowski, M., Jr. (1966) *J. Biol. Chem.* 241, 3955-3961
20. Abe, Y., Shirane, K., Yokosawa, H., Matsushita, H., Mitta, M., Kato, I., & Ishii, S. (1993) *J. Biol. Chem.* 268, 3525-3529
21. Yoshikawa, M., Kiyohara, T., Iwasaki, T., Ishii, Y., & Kimura, N. (1979) *Agric. Biol. Chem.* 43, 787-796
22. Norioka, S. & Ikenaka, T. (1983) *J. Biochem.* 93, 479-485
23. Terada, S., Fujimura, S., & Kimoto, E. (1994) *Biosci. Biotech. Biochem.* 58, 376-379
24. Joubert, F.J. & Dowdle, E.B. (1987) *Thromb. Haemost.* 57, 356-360
25. Kouzuma, Y., Suetake, M., Kimura, M., & Yamasaki, N. (1992) *Biosci. Biotech. Biochem.* 56, 1819-1824
26. Kim, S.H., Hara, S., Hase, S., Ikenaka, T., Toda, H., Kitamura, K., & Kaizuma, N. (1985) *J. Biochem.* 98, 435-448
27. Yamamoto, M., Hara, S., & Ikenaka, T. (1983) *J. Biochem.* 94, 849-863
28. Shibata, H., Hara, S., & Ikenaka, T. (1988) *J. Biochem.* 104, 537-543
29. Hattori, T., Yoshida, N., & Nakamura, K. (1989) *Plant. Mol. Biol.* 13, 563-573
30. Theerasilp, S., Hitotsuya, H., Nakajo, S., Nakaya, K., Nakamura, Y., & Kurihara, Y. (1989) *J. Biol. Chem.* 264, 6655-6659
31. Kortt, A.A., Strike, P.M., & De Jersey, J. (1989) *Eur. J. Biochem.* 181, 403-408
32. Morihara, K., Oka, T., & Tsuzuki, H. (1970) *Arch. Biochem. Biophys.* 129, 515-525
33. Kojima, S., Kumagai, I., & Miura, K. (1990) *Protein Eng.* 3, 527-530
34. Kyte, J. & Doolittle, R.F. (1982) *J. Mol. Biol.* 157, 105-132
35. Onesti, S., Brick, P., & Blow, D.M. (1991) *J. Mol. Biol.* 217, 153-176

## Supplemental Materials

### MATERIALS AND METHODS

*Materials*—The subtilisin inhibitors were purified as described previously (*10*). Trypsin treated with L-(1-tosylamido-2-phenyl)ethyl chloromethyl ketone (TPCK), *Achromobacter lyticus* lysyl endopeptidase, *Staphylococcus aureus* V8 protease, and all other chemicals were obtained from Wako Chem. Co. (Osaka).

Chemical and Proteolytic Digestions—The inhibitors were *S*-pyridylethylated by the method of Hermondson *et al.* (*11*). The *S*-pyridylethylated (PE)-proteins (3 mg each) were cleaved at room temperature with cyanogen bromide in 0.5 ml of 70 % formic acid (*12*). The PE-proteins (3 mg each) were digested separately with lysyl endopeptidase (E/S=1/100, by weight) in 20 mM Tris-HCl (pH 9.0) at 37 °C for 8 h, and with *S. aureus* V8 protease (E/S=1:50) in 0.1M $NH_4HCO_3$ at 37 °C for 8 h. A part of CLSI-III was also *S*-carboxymethylated (*13*) and digested with V8 protease as described above. Tryptic digestion was achieved with TPCK-trypsin (E/S=1:50) in 0.1M $NH_4HCO_3$ at 37 °C for 4 h. The digests were lyophilized, dissolved in 5 M urea-0.1% trifluoroacetic acid (TFA), and then fractionated by reverse-phase HPLC on a Biofine RPC-SC18 column (0.46 x 25 cm, Jasco) in 0.1% TFA with an appropriate gradient of acetonitrile.

*Determination of the Locations of Disulfide Bonds*—Intact CLSI-III (3 mg) was denatured in 0.3 ml of 0.1 N HCl at 40 °C for 1 h, and then the mixture was neutralized with 0.2 N NaOH. After TPCK-trypsin (60 μg) in 100 μl of 0.2 M Tris-HCl (pH 8.0) had been added, the mixture was incubated at 40 °C for 48 h. The resulting peptides were separated by reverse-phase HPLC on a C18 column as described above.

*Amino Acid Analysis and Sequence Determination*—The amino acid compositions of peptides were determined using an amino acid analysis system (Jasco) after hydrolysis with 0.25% phenol-2% thioglycolic acid-6M HCl (*14*) for 20 h at 110 °C in evacuated sealed tubes. The amino acid sequences of peptides were determined by manual Edman degradation using the DABITC/PITC double coupling method (*15*). Amino acid derivatives were identified by an HPLC method (*16*) on a Biofine RPC-SC18B column (4.6 x 250 mm; Jasco).

*Modification of Arginine Residues*—Phenylglyoxal treatment of the inhibitors was performed according to the method of Takahashi (*17*). To 200 μl of a protein solution (1 mg/ml) in 0.2 M *N*-ethylmorpholine-acetic acid buffer (pH 8.0) was added 200 μl of 0.3 % phenylglyoxal in the same buffer. After incubation at 25 °C for various periods, the remaining inhibitory activities were measured as described above. The inhibitor solution without phenylglyoxal was also incubated similarly as a control.

*Reactive-site Modification*—CLSI-III (1.1 mg, 50 nmol) was incubated in 100 μl of 0.1 M Na-phosphate buffer (pH 7.6) for 1 h at 30 °C with 68 mol% subtilisin BPN'(0.9 mg, 34 nmol). The reaction mixture was subjected to reverse-phase HPLC on a Biofine RPC-SC18 column, and elution was performed at the flow rate of 1 ml/min with a linear gradient of 0.1%TFA and acetonitrile (0-65% for 60 min). The inhibitor eluted at 49 min was recovered and *S*-pyridylethylated as described above. After excess reagent had been removed by dialysis against water, the inhibitor preparation was subjected to reverse-phase HPLC on the same column. The C-terminal amino acids were identified by carboxypeptidase digestion (*18*).

Table IIS. Amino acid compositions of V8 protease peptides derived from PE-CLSI-II[a].

| Amino acid | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 | V10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Asp | - | 5.0(5) | 1.9(2) | 4.1(4) | - | 2.3(2) | 2.0(2) | 2.0(2) | 1.1(1) | 3.0(3) |
| Thr | - | 0.9(1) | 1.9(2) | - | - | - | 2.6(3) | 1.1(1) | 2.7(3) | 2.0(2) |
| Ser | - | 3.0(4) | 2.0(2) | - | 1.6(2) | 3.2(4) | 2.7(3) | - | 1.6(2) | 1.0(1) |
| Glu | 1.0(1) | 1.3(1) | - | - | 1.1(1) | 2.3(2) | 4.6(5) | 1.3(1) | 2.2(2) | 1.2(1) |
| Pro | 1.0(1) | - | - | - | 2.1(2) | - | 2.0(2) | 1.2(1) | 2.6(3) | 4.0(4) |
| Gly | - | 1.1(1) | 1.1(1) | - | 1.5(2) | - | 4.5(5) | 1.7(2) | 4.0(4) | 3.9(4) |
| Ala | - | - | 1.3(1) | - | 1.0(1) | 1.4(1) | 1.0(1) | - | - | 2.4(2) |
| Val | - | - | - | 2.6(3) | - | 2.3(2) | 2.7(3) | 1.7(2) | 3.1(3) | 3.2(3) |
| Met | - | - | - | 1.2(1) | - | - | - | 0.8(1) | - | 1.0(1) |
| Ile | 0.9(1) | 1.2(1) | 1.1(1) | - | 0.9(1) | 1.2(1) | 1.8(2) | 1.1(1) | 3.1(3) | 2.0(2) |
| Leu | - | - | - | - | - | - | 1.2(1) | 1.8(2) | 2.3(2) | 2.2(2) |
| Tyr | - | 0.8(1) | - | - | - | - | 1.6(2) | - | - | 1.7(2) |
| Phe | 0.9(1) | - | 1.0(1) | - | 0.8(1) | 1.0(1) | 2.5(3) | 1.7(2) | 1.9(2) | - |
| His | - | - | - | - | - | - | 2.2(2) | 2.0(2) | - | - |
| Pec[b] | - | 0.9(1) | - | - | - | 1.0(1) | 1.8(2) | - | - | 0.8(1) |
| Trp | - | - | - | - | - | 0.6(1) | - | - | - | 0.5(1) |
| Lys | 2.0(2) | - | - | - | 1.0(1) | 1.1(1) | 3.0(3) | - | 1.0(1) | 1.0(1) |
| Arg | - | 0.9(1) | 0.9(1) | - | - | - | 0.9(1) | 1.7(2) | 0.9(1) | 1.6(2) |
| Total | 6 | 16 | 11 | 8 | 11 | 16 | 40 | 19 | 27 | 32 |
| Position | 82-87 | 144-159 | 179-189 | 1-8 | 9-19 | 88-103 | 104-143 | 160-178 | 35-81 | 20-51 |

[a]Values in parentheses are taken from the sequence data. [b]*S*-Pyridylethylcysteine.

Table IS. Amino acid compositions of lysyl endopeptidase peptides derived from PE-CLSI-II[a].

| Amino acid | K1 | K2 | K3 | K4 | K5 | K6 | K7T1 | K7T2 | K7T3 | K7T4 | K7T5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Asp | - | 1.1(1) | 1.0(1) | 2.1(2) | 1.0(1) | 3.1(3) | 3.0(3) | 2.0(2) | - | 4.0(4) | 3.7(4) |
| Thr | - | 1.0(1) | 1.7(2) | - | 2.6(3) | 1.6(2) | - | 1.0(1) | - | - | 3.1(3) |
| Ser | 2.7(3) | 1.8(2) | 1.2(1) | 1.0(1) | 1.0(1) | 0.8(1) | - | 3.1(4) | - | 2.1(2) | 2.1(2) |
| Glu | 1.1(1) | - | 4.2(4) | 2.3(2) | 2.3(2) | 3.2(3) | 1.1(1) | 1.0(1) | - | 1.3(1) | 1.3(1) |
| Pro | 1.1(1) | 0.8(1) | 1.0(1) | - | 1.8(2) | 2.2(2) | - | - | 2.6(3) | 2.0(2) | 0.6(1) |
| Gly | - | - | 5.0(5) | - | 3.3(3) | 2.2(2) | 2.0(2) | 1.3(1) | 4.1(4) | 2.0(2) | 1.9(2) |
| Ala | 1.0(1) | - | 1.3(1) | - | - | - | - | - | 1.2(1) | 1.0(1) | 1.3(1) |
| Val | - | - | 1.6(2) | 1.9(2) | 0.8(1) | 3.7(4) | - | 1.1(1) | 0.9(1) | 3.3(3) | 2.0(2) |
| Met | - | - | - | - | - | - | - | - | 0.9(1) | 1.1(1) | 1.2(1) |
| Ile | - | - | 1.9(2) | 1.2(1) | 2.8(3) | 0.8(1) | - | 0.9(1) | 2.2(2) | 1.2(1) | 2.0(2) |
| Leu | - | - | - | - | 2.1(2) | 1.0(1) | - | 1.0(1) | - | - | 2.1(2) |
| Tyr | - | 1.8(2) | - | - | - | 0.9(1) | 1.1(1) | - | - | 2.0(2) | - |
| Phe | - | - | 1.7(2) | 1.0(1) | 2.0(2) | 1.1(1) | - | 0.9(1) | - | 1.0(1) | 3.0(3) |
| His | - | - | 1.9(2) | - | - | - | - | - | - | - | 2.4(2) |
| Pec[b] | 0.8(1) | - | 0.8(1) | - | - | 0.9(1) | - | 1.8(2) | - | - | - |
| Trp | - | - | - | 0.8(1) | - | - | - | - | 1.1(1) | - | - |
| Lys | 1.9(2) | 0.9(1) | 0.9(1) | 1.0(1) | 0.8(1) | 0.9(1) | - | - | - | 0.7(1) | - |
| Arg | - | 1.0(1) | - | - | 0.9(1) | 1.1(1) | 1.0(1) | 0.9(1) | 1.0(1) | - | 0.7(1) |
| Total | 9 | 9 | 25 | 11 | 22 | 24 | 8 | 16 | 14 | 21 | 27 |
| Position | 85-93 | 130-138 | 105-129 | 94-104 | 63-84 | 39-62 | 155-162 | 139-154 | 22-35 | 1-21 | 164-190 |

[a]Values in parentheses are taken from the sequence data. [b]*S*-Pyridylethylcysteine.

Table IIIS. Amino acid compositions of CNBr peptides derived from PE-CLSI-II[a].

| Amino acid | M1 | M2 | M3 | M4 |
|---|---|---|---|---|
| Asp | 3.0(3) | 3.2(3) | 1.1(1) | 12.7(14) |
| Thr | - | 0.9(1) | - | 9.0(10) |
| Ser | - | 0.9(1) | 1.6(2) | 12.6(13) |
| Glu | - | - | 1.1(1) | 18.7(15) |
| Pro | - | - | 1.3(2) | 11.3(11) |
| Gly | - | - | 2.3(2) | 17.4(17) |
| Ala | - | 1.0(1) | 1.2(1) | 4.2( 4) |
| Val | 2.6(3) | - | - | 11.8(13) |
| Met[b] | 0.9(1) | - | 0.6(1) | 0.6( 1) |
| Ile | - | - | 2.0(2) | 10.4(10) |
| Leu | - | - | - | 7.7( 7) |
| Tyr | - | - | 1.8(2) | 3.6( 4) |
| Phe | - | 1.0(1) | 0.9(1) | 8.9( 9) |
| His | - | - | - | 4.0( 4) |
| Pec[c] | - | - | - | 4.4( 5) |
| Trp | - | - | - | 1.5( 2) |
| Lys | - | - | 0.9(1) | 8.2( 8) |
| Arg | - | - | - | 7.0( 7) |
| Total | 7 | 7 | 16 | 154 |
| Position | 1-7 | 178-190 | 8-23 | 24-177 |

[a]Values in parentheses are taken from the sequence data. [b]Determined as homoserine plus homoserine lactone. [c]*S*-Pyridylethylcysteine.

Table IVS. Amino acid compositions of CNBr peptides derived from PE-CLSI-III, and peptides derived on tryptic digestion of M4 in Fig. 5S[a].

| Amino acid | M1 | M2 | M3 | M4K1 | M4K2 | M4K3 | M4K4 | M4K5 | M4K6 | M4K7 | M4K8 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Asp | 2.2(2) | 3.2(3) | 1.1(1) | - | 1.0(1) | - | 1.0(1) | 1.9(2) | 5.8(6) | 1.0(1) | 3.4(3) |
| Thr | 0.8(1) | - | - | - | 0.9(1) | - | 1.7(2) | - | 1.7(2) | 2.6(3) | 1.8(2) |
| Ser | 1.0(1) | - | 1.6(2) | 2.8(3) | 1.7(2) | - | 0.9(1) | 0.9(1) | 3.2(4) | 1.0(1) | 0.9(1) |
| Glu | - | - | 1.0(1) | 1.1(1) | - | - | 3.9(4) | 2.0(2) | 3.1(3) | 2.0(2) | 3.1(3) |
| Pro | - | - | 1.6(2) | 0.9(1) | 1.0(1) | 3.0(3) | 1.1(1) | - | 1.0(1) | 2.0(2) | 2.1(2) |
| Gly | - | - | 1.9(2) | - | - | 4.1(4) | 5.0(5) | - | 3.0(3) | 3.1(3) | 2.0(2) |
| Ala | 1.1(1) | - | 1.0(1) | 1.0(1) | - | 1.9(2) | 1.1(1) | - | - | - | - |
| Val | - | 2.4(3) | - | - | - | 1.0(1) | 1.9(2) | 1.9(2) | 2.5(3) | 1.0(1) | 3.8(4) |
| Met[b] | - | 0.5(1) | 0.8(1) | - | - | - | - | - | 0.4(1) | - | - |
| Ile | - | - | 1.9(2) | - | - | 1.0(1) | 1.9(2) | 0.9(1) | 1.7(2) | 3.2(3) | 0.7(1) |
| Leu | - | - | - | - | - | 1.0(1) | - | - | 2.9(3) | 1.9(2) | 1.0(1) |
| Tyr | - | - | 2.0(2) | - | 1.9(2) | - | - | - | 0.8(1) | - | 1.0(1) |
| Phe | 1.0(1) | - | 1.0(1) | - | - | - | 1.8(2) | 1.0(1) | 2.9(3) | 1.8(2) | 1.0(1) |
| His | - | - | - | - | - | - | 1.9(2) | - | 1.7(2) | - | - |
| Pec[c] | - | - | - | 0.6(1) | - | - | 0.8(1) | - | 1.6(2) | - | 1.1(1) |
| Trp | - | - | - | - | - | 0.9(1) | - | 0.8(1) | - | - | - |
| Lys | - | - | 1.0(1) | 2.0(2) | 1.0(1) | 1.0(1) | 1.0(1) | 1.1(1) | - | 1.0(1) | 1.0(1) |
| Arg | - | - | - | - | 1.0(1) | 1.0(1) | - | - | 3.1(3) | 0.9(1) | 1.0(1) |
| Total | 6 | 7 | 16 | 9 | 9 | 15 | 25 | 11 | 39 | 22 | 24 |
| Position | 178-183 | 1-7 | 8-23 | 85-93 | 130-138 | 24-38 | 105-129 | 94-104 | 139-177 | 63-84 | 39-62 |

[a]Values in parentheses are taken from the sequence data. [b]Determined as homoserine plus homoserine lactone. [c]S-Pyridylethylcysteine.

Table VIS. Amino acid compositions of tryptic peptides derived from non-reduced CLSI-III[a].

| Amino acid | NT1 | NT2 | NT3 |
|---|---|---|---|
| Asp | 2.1(2) | 3.2(3) | 1.4(1) |
| Thr | 0.8(1) | 1.1(1) | 2.0(2) |
| Ser | 3.4(4) | 3.6(4) | 1.0(1) |
| Glu | 1.3(1) | 4.0(4) | 4.3(4) |
| Pro | - | 2.9(3) | 1.0(1) |
| Gly | - | 1.7(2) | 5.2(5) |
| Ala | - | 1.2(1) | 1.0(1) |
| Cys[b] | 1.5(2) | 1.4(2) | 0.7(1) |
| Val | 1.1(1) | 4.1(4) | 2.0(2) |
| Ile | 1.0(1) | 1.1(1) | 1.6(2) |
| Leu | 1.0(1) | 1.2(1) | - |
| Tyr | - | 1.1(1) | - |
| Phe | 1.0(1) | 0.7(1) | 2.4(2) |
| His | - | - | 1.8(2) |
| Lys | - | 2.5(3) | 1.0(1) |
| Arg | 1.0(1) | - | - |
| Total | 15 | 31 | 25 |
| Position | 139-154 | 41-62<br>85-93 | 105-129 |
| Position of disulfide bridge | $Cys^{142}$-$Cys^{149}$ | $Cys^{44}$-$Cys^{88}$ | $Cys^{105}$ (free SH) |

[a]Values in parentheses are taken from the sequence data. [b]Determined as cysteine.

Table VS. Amino acid compositions of V8 protease peptides derived from CM-CLSI-III[a].

| Amino acid | V1 | V2 | V3 | V4 | V5 | V6 | V7 | V8 | V9 |
|---|---|---|---|---|---|---|---|---|---|
| Cmc[b] | - | 0.8(1) | - | - | 0.7(1) | 1.2(2) | - | 0.7(1) | - |
| Asp | - | 4.9(5) | 3.8(4) | - | 1.9(2) | 2.0(2) | 1.1(1) | 3.0(3) | 2.9(3) |
| Thr | - | 1.0(1) | - | - | - | 2.9(3) | 2.5(3) | 1.9(2) | 1.8(2) |
| Ser | - | 3.4(4) | - | 1.5(2) | 3.3(4) | 2.9(3) | 2.1(2) | 1.1(1) | 1.1(1) |
| Glu | 1.1(1) | 1.2(1) | - | 1.1(1) | 2.2(2) | 5.0(5) | 2.1(2) | 2.2(2) | 1.1(1) |
| Pro | 1.1(1) | - | - | 1.9(2) | - | 2.0(2) | 2.6(3) | 4.0(4) | 1.0(1) |
| Gly | - | 1.3(1) | - | 1.9(2) | - | 4.6(5) | 4.1(4) | 5.3(5) | 2.1(2) |
| Ala | - | - | - | 0.9(1) | 1.1(1) | 1.0(1) | - | 2.1(2) | 1.0(1) |
| Val | - | - | 3.0(3) | - | 1.8(2) | 2.9(3) | 2.5(3) | 3.0(3) | 1.6(2) |
| Met | - | - | 1.1(1) | - | - | - | - | 1.0(1) | 0.9(1) |
| Ile | 1.0(1) | 1.1(1) | - | 1.0(1) | 1.0(1) | 2.1(2) | 2.9(3) | 1.9(2) | 1.0(1) |
| Leu | - | - | - | - | - | 1.2(1) | 1.9(2) | 2.0(2) | 1.8(2) |
| Tyr | - | 1.0(1) | - | - | - | 1.7(2) | - | 2.7(3) | - |
| Phe | 1.0(1) | - | - | 1.0(1) | 0.9(1) | 2.7(3) | 1.8(2) | - | 2.6(3) |
| His | - | - | - | - | - | 1.5(2) | - | - | 1.9(2) |
| Trp | - | - | - | - | 0.7(1) | - | - | 0.7(1) | - |
| Lys | 1.8(2) | - | - | 1.1(1) | 1.1(1) | 2.6(3) | 1.0(1) | 1.0(1) | - |
| Arg | - | 0.9(1) | - | - | - | 1.1(1) | 0.9(1) | 1.8(2) | 1.8(2) |
| Total | 6 | 16 | 8 | 11 | 16 | 40 | 27 | 35 | 24 |
| Position | 82-87 | 144-159 | 1-8 | 9-19 | 88-103 | 104-143 | 55-81 | 20-54 | 160-183 |

[a]Values in parentheses are taken from the sequence data. [b]S-Carboxymethylcysteine.

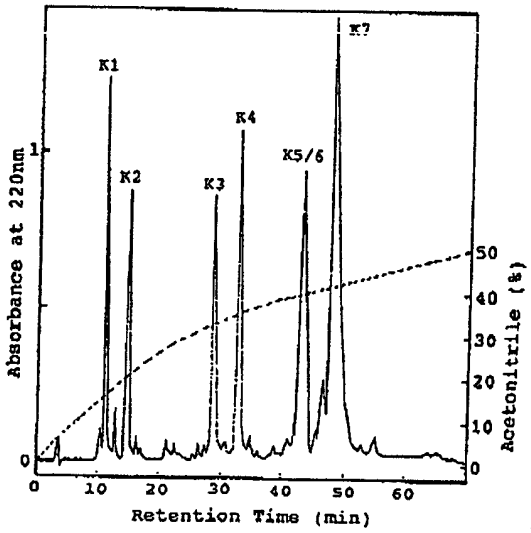


Fig. 1S. Fractionation of a lysyl endopeptidase digest of PE-CLSI-II by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA.

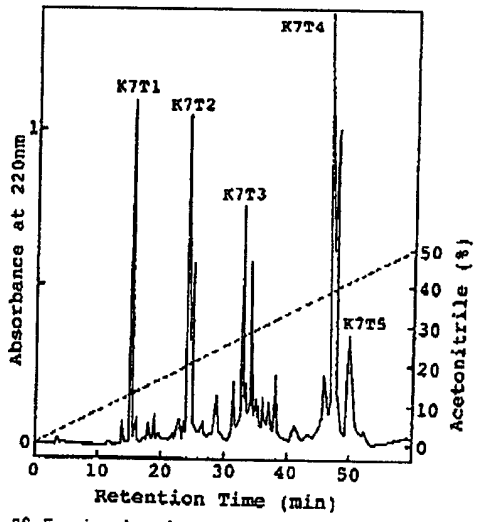


Fig. 2S. Fractionation of a tryptic digest of peptide K7 in Fig. 1S by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA.

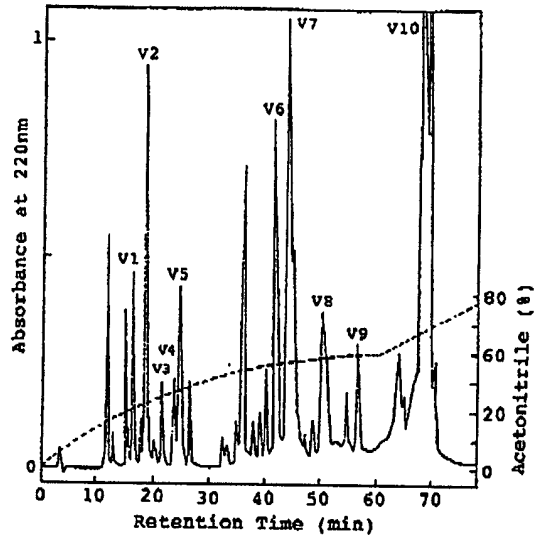


Fig. 3S. Fractionation of a V8 protease digest of PE-CLSI-II by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA.

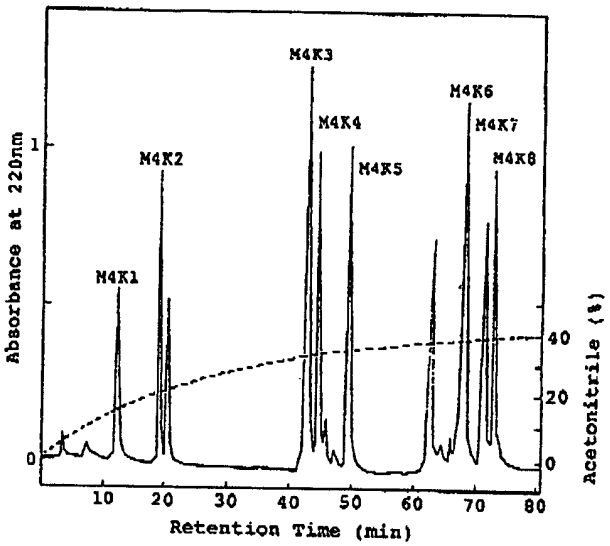


Fig. 6S. Fractionation of a tryptic digest of peptide M4 in Fig. 5S by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA.

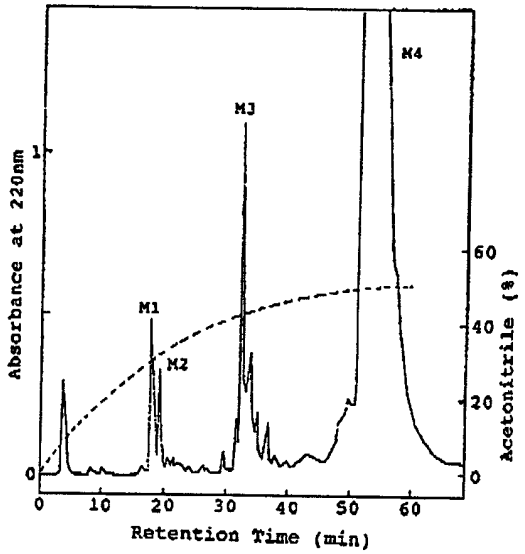


Fig. 4S. Fractionation of cyanogen bromide peptides of PE-CLSI-II by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA.

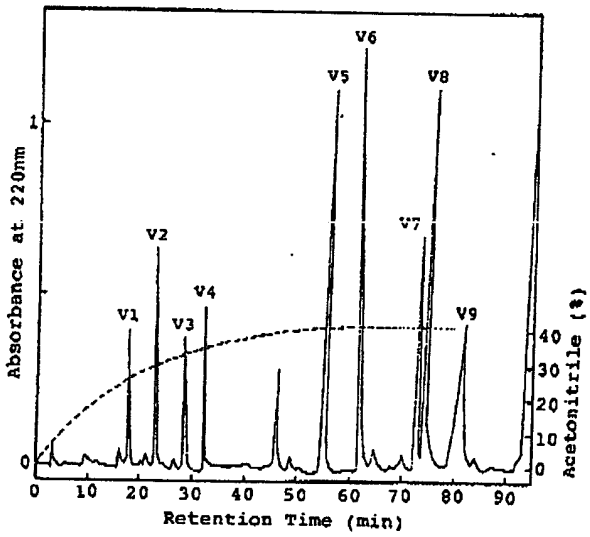


Fig. 7S. Fractionation of a V8 protease digest of CM-CLSI-III by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA.

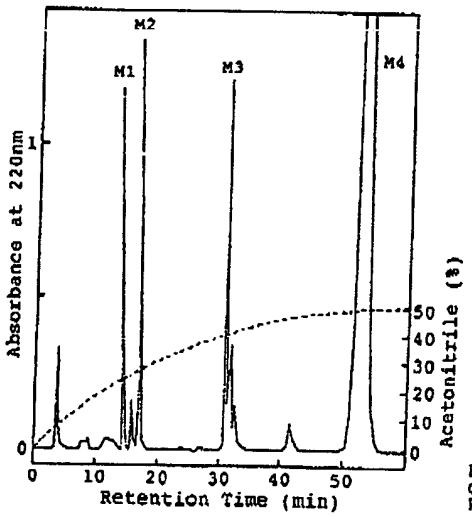


Fig. 5S. Fractionation of cyanogen bromide peptides of PE-CLSI-III by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA.

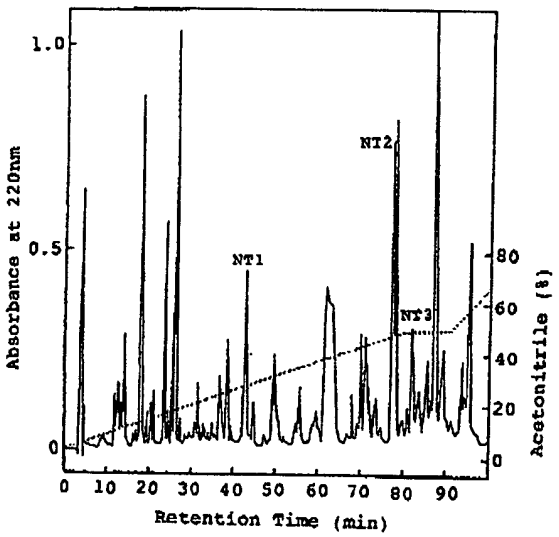


Fig. 8S. Separation of a tryptic digest of non-reduced CLSI-III by reverse-phase HPLC on a column of Biofine RPC-SC18 with a gradient of acetonitrile in 0.1 % TFA. Cys-containing peptides were detected by amino acid analyses.

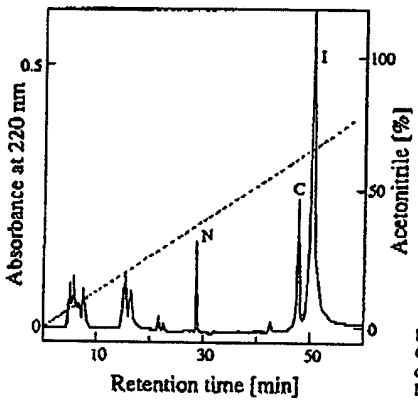


Fig. 9S. Separation of fragments derived from the modified CLSI-III and the intact inhibitor by reverse-phase HPLC on a column of Biofine RPC-SC18. N, N-fragment; C, C-fragment; I, intact inhibitor.

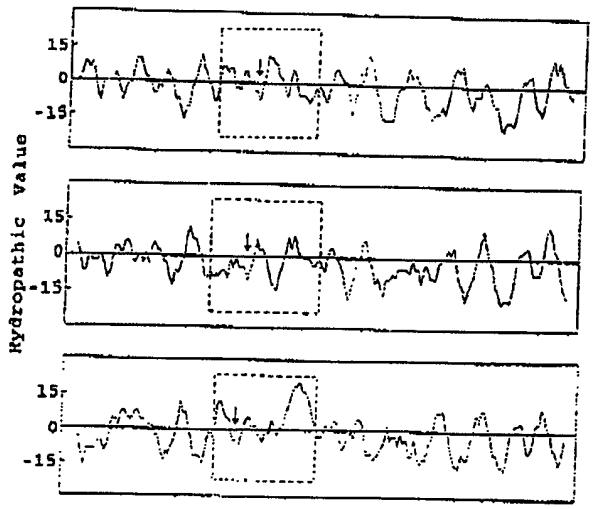


Fig. 10S. Comparison of the hydropathy profiles of CLSI-III (top), *E. caffra* trypsin inhibitor DE-3 (middle), and soybean trypsin inhibitor (bottom). Hydropathy was calculated by the method of Kyte and Doolittle (*34*). Dissimilar regions are boxed by broken lines. Arrows indicate the positions of the putative reactive sites.

Vol. 31, No. 1, September 1993 BIOCHEMISTRY *and* MOLECULAR BIOLOGY INTERNATIONAL
Pages 73-81

# COMPARISON OF AMINO ACID SEQUENCES OF THE TRYPSIN INHIBITORS FROM TARO (*COLOCASIA ESCULENTA*), GIANT TARO (*ALOCASIA MACRORRHIZA*) AND GIANT SWAMP TARO (*CYRTOSPERMA CHAMISSONIS*)

Liangcai Peng[1], J. Howard Bradbury[1], Brendon C. Hammer[2] and Denis C. Shaw[3]+

[1]Division of Botany & Zoology, Australian National University, Canberra, ACT 0200, Australia.
[2]Department of Chemistry, Australian National University, Canberra, ACT 0200, Australia.
[3]Protein Biochemistry Group, John Curtin School of Medical Research, Australian National University, Canberra, ACT 2601, Australia.

Received August 17, 1993
Received after revision, August 19, 1993

## SUMMARY

The amino acid sequences of the trypsin inhibitors from taro *Colocasia esculenta* var. *esculenta* and giant swamp taro *Cyrtosperma chamissonis* have been determined and are compared with the protein sequence of the trypsin/chymotrypsin inhibitor from giant taro *Alocasia macrorrhiza*. Both inhibitors display polymorphism and there is evidence of two components in the giant swamp taro. The positional identity between the proteins is highest at 73-75% for the comparison of the giant taro (GT) with the polymorphic forms of the taro (T) inhibitors and lowest at 56-58% for the pairs of taro and giant swamp taro (GST) proteins. The comparisons show that the inhibitors from T and GT are more related to each other than to GST, which supports their taxonomic classification into different tribes. Location of the $P_1$ site for the trypsin inhibitors of aroids is different from that of other Kunitz-type inhibitors and could be at Leu56.

## INTRODUCTION

Many protease inhibitors from the plant families Leguminosae, Gramineae and Solanaceae have been studied over nearly fifty years (1,2) but there has been much less work done on those from the Araceae plant family. The latter contains a group of related plants called the edible aroids, the most important of which (taro *Colocasia* and taro *Xanthosoma*) form the staple diet of 200 million people living in the tropics (3-5). Taro *Xanthosoma sagittifolium* is essentially free of trypsin and chymotrypsin inhibitors (6), but taro *Colocasia esculenta* var. *esculenta* (7,8) and a closely related plant, taro *Colocasia esculenta* var. *antiquorum* (9-12) contain appreciable amounts of trypsin inhibitors. There are three other minor edible aroids of some importance in

+To whom all correspondence should be addressed.

1039-9712/93/130073-09$05.00/0

the Pacific. Of these elephant foot yam (*Amorphophallus campanulatus* ) has virtually no trypsin inhibitor (13), giant swamp taro (*Cyrtosperma chamissonis*) contains appreciable amounts and giant taro (*Alocasia macrorrhiza*) large amounts of trypsin inhibitor (8-14). The proteinase inhibitor from giant taro inhibits both trypsin and chymotrypsin (7,8,13) and its amino acid sequence has been determined (15). The trypsin inhibitors from taro, giant taro and giant swamp taro have similar structures. They each consist of two monomers (shown to be identical for giant taro (15)) with homology to the Kunitz-type inhibitor from soybean. The inhibitor dimers of molecular mass ca. 40 kDa from the taro family bind two molecules of trypsin, ie. one per monomer (7,15).

Because the protease inhibitor from giant taro is the only one of the aroid proteins to also inhibit chymotrypsin, it is of interest to compare the amino acid sequences of taro and giant swamp taro with that of giant taro. Accordingly, in this paper we report the amino acid sequences of the trypsin inhibitors from taro (*C. esculenta* var. *esculenta*) and giant swamp taro and compare them with the known sequence of giant taro (15).

## MATERIALS AND METHODS

**Materials:** Corms of taro (*C. esculenta* var. *esculenta*) cv. "Samoa Green" were obtained from Fiji and giant swamp taro, cv. "Ikaraoi Green" was from Kiribati. Corms were processed and trypsin inhibitors obtained as described previously (7). Sequencing grade enzymes were obtained from Boehringer Mannheim Australia.

**Protein sequencing:** Enzymic and chemical cleavage of the inhibitors and the separation of peptide fragments were as described previously (15). Protein/peptide sequencing was on an Applied Biosystems 477A sequencer with on-line 120A PTH-Analyser. The major and minor components of the giant swamp taro inhibitor were separated on a reverse phase $C_8$ column using the Pharmacia SMART System.

## RESULTS

**Amino acid sequences:** An overlapping sequence of the taro (T) trypsin inhibitor was established using tryptic, Glu-C, Asp-N and CNBr cleavages with separation of the fragments by HPLC. The sequence of giant swamp taro (GST) trypsin inhibitor was determined from tryptic, chymotryptic and CNBr cleavages. These sequences together with that of the trypsin/chymotrypsin inhibitor from giant taro (GT), previously obtained by this laboratory (15), are shown in Fig. 1. A limited number of gaps have been introduced to maximise positional identities and similarities, the numbering shown is that for the giant taro (GT) inhibitor. It should be noted that in our previous publication (7), the *N*-terminal sequences for taro (*Colocasia esculenta* ) and giant swamp taro (*Cyrtosperma chamissonis*) were transposed, the correct assignment being as in Fig. 1.

```
                    10        20        30        40        50        60
                    |         |         |         |         |         |
(GTa)    TNPVLDVDGNELQRGQLYYATSV-MRPGGGLTLAAPKGS--CPLNVAQA-PFDEYSGRPLAFFPENADD
(GTb)                                  A                            K

(Ta)     ANPVLDVDGNELRRGNRYYAISL-RRPNSGLTLAARS-NAPCPLNVAQAAPRDY--GRPLAFFPENADD
(Tb)        I                   E    V                          E

(GSTa)   ADAVLDLEGNELQRGQLYYAQAIFRGSNSGLRLSAFSGS--CPLYVTKA-SFGDPTGEPLAFFPENPNE
(GSTb)                    N N    E    Q                    S G P N  S R   L

                    70        80        90        100       110       120       130
                    |         |         |         |         |         |         |
(GTa)    DTVQEGSTLYIMFPEPTRCPQSTVWTFDREAGFVTTGGTTSKAIGPHNSRFAIRKAGDASSQPRDYQI

(Ta)     DTVQEGSTLNIMFPEPTECRESTVWTLDRETGVVTTGGTSSSEIGPHNERFAIRRAGDASSK-REYQI
(Tb)                      Q  R         G  R

(GSTa)   NIIREGSTLYIVFAEPTRCGESTAWTLDSRTSYVTTGGSSSTAVGPHSSRFAIYRAGGTGFN--YYQF
(GSTb)       RA     AM PQ  E            G                          E AS

                    140       150       160       170       180
                    |         |         |         |         |
(GTa)    EVCPCSTGVERPSCRMGCLGTLGLAEGGK-NVLLNINNESPHTIRFVKVKEG

(Ta)     EVCPCSIGVSRPSCRMACVGSLGLTEDEA-NLLLNINNERPHTVRFVKVQEEL
(Tb)     T     L         L                              E         I

(GSTa)   QVCPCSIGSGRPSCRMGCTGGLGLASGQNENV-LTTNSNSPHNVIFVKVKEG
(GSTb)                    A         RE
```

Fig. 1. Amino acid sequences of the trypsin inhibitors from the aroids , taro (T) and giant swamp taro (GST) compared with the trypsin/chymotrypsin inhibitor from giant taro (GT). The numbering is for GT with gaps, --, inserted to facilitate alignment and maximise positional identities. Polymorphism observed in the protein sequencing is indicated by the use of two lines, eg. (Ta) and (Tb), where alternate amino acids were detected at a position one is shown in line (Ta) and the other in line (Tb), see text. The underlining in line (GSTa) is to indicate where an amino acid is conserved in all three aroids and polymorphism not observed in protein sequencing.

**Polymorphism:** In each case there was evidence of polymorphism. There were only two positions in the giant taro inhibitor sequence where polymorphism was detected; a Met or Ala occurred in equal abundance at 24, while at position 50 there was Glu or Lys in a ratio of 4:1 in the sample subjected to protein sequencing. The successful cloning preferentially isolated the Met24 and Lys50 forms (A. Matthews et al., personal communication). In the taro inhibitor there were thirteen sites of polymorphism detected in the protein sequencing. These are shown on Fig. 1 by having in line (Ta) the complete sequence with one of the possible amino acid residues at each site of polymorphism, while in line (Tb) only the alternate residues at those sites are indicated (eg. Ile or Val at position 4), the rest of the sequence being as in (Ta). The polymorphisms shown occurred in approximately equal proportion, however it is not possible to say from the protein sequencing of the various isolated fractions which combination of amino acids occur in any one molecule of inhibitor. The previously reported (7) isoelectric focussing profile of the taro inhibitor sample showed five discreet bands plus a diffuse area. It could be argued that the sequencing results are consistent with the observed micro heterogeneity. Given that seven of the substitutions involve charge differences, (five involving a change between an uncharged and a negative charge, one between an uncharged and a positive charge and one between a negative and a positive charge) and that the migrating species would be homo- and heterodimers (6) there are unlikely to be many different monomer species and there would be internal compensation of the charge differences.

Since the completion of the protein sequencing we have become aware of some work on the cloning of taro (*Colocasia esculenta* Schott) storage proteins (16). Comparison of the amino acid sequence of the protein and that deduced from the two clones (named 5.1 and 10.1) for the 25 kDa protein clearly indicates that the supposed storage protein is in fact the taro trypsin inhibitor. In Table 2 are listed the sites of polymorphism observed in both the protein and cDNA sequencing of the taro trypsin inhibitor. For completeness Table 2 includes those positions where, while there is no change in the amino acid, there is change in the third base of the codon for the amino acid. Also noted are those positions where there is difference between the protein and nucleotide derived sequences although no difference was observed within the two sets of data. Attention is drawn to those positions in the taro sequences where polymorphism has been observed also in our protein sequencing of the giant swamp taro (GST) inhibitor and to where the protein sequencing indicates some conservation of an amino acid at a position in the chain. Positions where gaps have been introduced in either the protein or cDNA sequences in order to facilitate the alignment of the sequences are listed. The numbering used in Table 2 is as in Fig. 1, ie. that for the giant taro (GT) inhibitor. In the alignment of GT, T and GST sequences in Fig. 1 the taro sequence has an additional residue between positions 40-50; this is numbered 46A in the protein and it can be noted that a gap is introduced in the nucleotide sequence, while there is a compensating insertion in that sequence at 48A to maximise the alignment. In the region of

positions 128-130 the two clones are different and 10.1 has six extra bases (two amino acids). This occurs where, in the alignment of the inhibitors in Fig. 1, gaps have been introduced and may correspond to the junction of exons. As noted above, the assignment of amino acids from the protein sequencing to lines Ta or Tb was purely arbitrary. Within a clone there is indication of a particular sequence of amino acids, however only two clones have been sequenced whereas at least five protein species have been observed. Therefore until further clones, which relate to variations of the protein sequence, have been studied the assignment of amino acids must remain arbitrary. A reallocation of amino acids to lines Ta and Tb on the basis of the limited number of clones would not be consistent. Some differences between the two sets of data could be expected, due to possible differences in the variety of taro used in each case.

For the giant swamp taro (GST) the meaning of lines (GSTa and b) is somewhat different from lines (Ta and b) of taro. The isoelectric focussing profile of the GST inhibitor had shown a small number of bands within a broad diffuse area and during the sequencing it became apparent that the sample, as isolated from a trypsin affinity column, consisted of a major and a minor component which could be partially separated by reverse phase HPLC. From the heights of the peaks of the trial HPLC, the two components occurred in a ratio of ca. 4:1 and they both gave the same *N*-terminal sequence to twenty residues. The mixture was subjected to enzymic digestion and the resultant peptides separated by HPLC and sequenced. The data were interpreted being mindful that there were two components with polymorphic variation expected. Within line (GSTa) is shown the sequence of a tryptic peptide which occurred as a major peak and resulted from hydrolysis after Lys45 and Arg69. In line (GSTb) are those residues which differed in that region. The alternate sequence came from a smaller peak (ca. one quarter the height of the major peak) which eluted off the $C_{18}$ column at a slightly lower %B (15). Consistent with normal observations trypsin did not cleave the Arg-Pro bond at 54-55 and it appeared that there was preferential cleavage by the endoprotease after an Arg-Arg sequence in this peptide, which extended that form to position 70. It is feasible that the sequence from 46 to 69 in line (GSTa) was derived from the major component and that in line (GSTb), residues 46-70 were derived from the minor component. While work is in progress to test this interpretation, it is clear that in that region there are two substantially different sequences: the differences occur as a set and do not mean that there is multiple polymorphic variation in this region. Hence while there is a major difference between the two forms of inhibitor as isolated from the giant swamp taro and this is shown in the region 46-70 of lines (GSTa and b), the other sites of difference could be in either component and it is noted that some occur at positions where polymorphism has been observed in taro.

**Positional identities:** In Fig. 1 the underlining in line (GSTa) indicates those residues which are common to all forms of the inhibitors from the three aroid species. There are 80 such residues,

TABLE 1. Postitional Identities and Similarities in Trypsin Inhibitors

For each pair of inhibitors the Percent of Postitional Identities are listed, also tabulated are the Quality and Percent Similarity between pairs. The data were derived by application of the GCG program 'Gap' which uses the algorithm of Needleman and Wunsch (17). The inhibitors are as in Fig. 1: GTa, giant taro; GTb, giant taro polymorphic variant; Ta, taro; Tb, taro variant; GSTa, giant swamp taro; GSTb, giant swamp taro variant (see text).

Quality* (top) and Percent Similarity (bottom)

Percent Positional Identity

| Inhibitor | GTa | GTb | Ta | Tb | GSTa | GSTb |
|---|---|---|---|---|---|---|
| GTa | Quality | 273.3 | 214.5 | 208.8 | 182.2 | 178.5 |
| | % Similarity | 98.9 | 84.1 | 83.0 | 75.8 | 75.8 |
| GTb | 98.9 | | 213.3 | 207.1 | 180.5 | 176.8 |
| | | | 83.5 | 82.4 | 74.7 | 74.7 |
| Ta | 75.3 | 75.3 | | 261.1 | 164.0 | 164.0 |
| | | | | 96.2 | 71.8 | 72.4 |
| Tb | 73.1 | 72.5 | 92.4 | | 156.5 | 155.7 |
| | | | | | 70.2 | 70.7 |
| GSTa | 58.2 | 58.2 | 56.4 | 52.5 | | 248.1 |
| | | | | | | 92.4 |
| GSTb | 56.0 | 56.0 | 55.2 | 52.5 | 86.9 | |

*Quality = ComparativeValue (CmpVal)$_{AA}$ x Total$_{AA}$ + CmpVal$_{AB}$ x Total$_{AB}$ + CmpVal$_{AC}$ x Total$_{AC}$ ..... + CmpVal$_{ZZ}$ x Total$_{ZZ}$ - (GapWeight x GapNumber) - (GapLengthWeight x Total LengthOfGaps). GapWeight used = 3.0, GapLengthWeight = 0.10.

which is 43.5% over a total of 184. The percent positional identity, quality and percent similarity for each pair of inhibitors are shown in Table 1. The pairwise comparisons were performed with the Wisconsin GCG program 'Gap' which uses the algorithm of Needleman and Wunsch (17) and gives values for the quality and percent similarity using a symbol comparison table, with matches equal to 1.5 and mismatches based on the evolutionary distance between the amino acids (18). The values in the line or row for say Ta refer to those for the sequence shown as line (Ta) in Fig. 1, while those in Tb are derived from line (Tb). Because of the polymorphism and not knowing which of the alternate amino acids actually occurs in a particular molecule of inhibitor, the values in each case should be viewed as giving a range for the species. There is 92.4% positional identity between the sequences shown in lines (Ta and b) and when GTa is compared with each of them there is a range of from 73.1-75.3%. The positional identity between the sequences

TABLE 2. Polymorphisms in Taro Trypsin Inhibitor

The amino acids which occur at each site of polymorphism in the taro trypsin inhibitor are shown in columns Ta and Tb, which are from the protein sequences found in Fig. 1, and 5.1 and 10.1, which refer to the clones isolated by Hirai et al. (16), their nucleotide sequence being translated to amino acids. The numbering is as in Fig. 1, i.e. for the giant taro (GT) sequence.

| Position | Ta | Tb | 5.1 | 10.1 | Notes | Position | Ta | Tb | 5.1 | 10.1 | Notes |
|---|---|---|---|---|---|---|---|---|---|---|---|
| -2 | | | V | A | | 96 | T | R | T | R | c |
| -1 | | | A | A | a* | 97 | G | G | D | G | |
| 1 | A | A | S | S | b | 98 | V | V | V | V | a |
| 4 | V | I | V | I | | 102 | G | G | G | G | a,d |
| 6 | D | D | D | D | a | 108 | E | E | A | A | a,h |
| 10 | N | N | N | D | | 111 | P | P | P | P | a,d |
| 16 | N | N | N | H | c | 112 | H | H | R | H | d |
| 23 | L | E | L | E | | 114 | E | E | S | S | b,i |
| 25 | R | R | S | R | c | 117 | A | A | T | S | d |
| 27 | N | V | N | V | | 121 | A | A | T | A | d |
| 34 | A | A | V | A | d | 127 | K | K | K | E | |
| 39 | C | C | C | C | a | 128 | - | - | - | R | e |
| 44 | A | A | D | A | c | 128A | - | - | G | - | e |
| 46 | A | A | A | T | c | 128B | - | - | E | E | e |
| 46A | A | A | - | - | e | 129 | R | R | - | R | e |
| 47 | P | P | P | S | | 130 | E | E | - | K | e |
| 48 | R | R | S | S | b | 134 | E | T | E | E | b |
| 48A | - | - | K | N | e | 140 | I | L | I | N | |
| 49 | D | E | D | D | a,c | 143 | S | S | S | P | |
| 53 | G | G | G | G | a | 145 | P | P | A | P | d |
| 63 | A | A | V | A | | 146 | S | S | P | S | d |
| 64 | D | D | D | E | | 147 | C | C | C | C | a,d |
| 68 | V | V | V | V | a | 149 | M | L | L | M | j |
| 72 | S | S | S | N | d | 151 | C | C | C | C | a,d |
| 75 | N | N | Y | N | f | 153 | G | G | G | G | a,d |
| 80 | E | Q | E | E | b | 162 | A | A | A | G | |
| 81 | P | P | P | P | a,d | 175 | T | E | T | A | |
| 82 | T | T | S | T | d | 176 | V | V | V | I | k |
| 83 | E | R | E | E | b,c | 182 | Q | Q | K | K | b |
| 86 | E | E | E | A | | 185 | L | I | L | L | b,l |
| 91 | T | T | T | T | a,d | 186 | - | - | A | P | |
| 93 | D | G | G | D | g | | | | | | |

*a, change in third base of codon; b, no difference in nucleotide sequence; c, site of polymorphism in GST; d, conserved amino acid in protein alignment; e, gap in protein sequence or clone to align sequence; f, Y in GST; g, D in GT and GST; h, A in GT and GST; i, S in GT and GST; j, M in GT and GST; k, I in GT; l, *C*-terminus.

shown in lines (GSTa and b) for the giant swamp taro inhibitors is lower at 86.9% and the comparisons with GTa gives a range of 56.0-58.2%. There is clearly more positional identity between GT and T than GT and GST. The least similar are T and GST, with an overall range of only 52.5-56.4% positional identity. The data for quality and percent similarity show the same trend with the best values for GT compared with T having an overall range of 207.1-214.5 for quality and 82.4-84.1% similarity. The least similar are T to GST, the overall range being 155.7-164.0 for quality and 70.2-72.4% similarity. The inhibitors from T and GT appear to be more related to one another than to GST, which supports the taxonomic classification of T and GT in the tribe Colocasiodeae and GST in the Lasioideae (19).

## DISCUSSION

We know from our previous studies (7) that the proteins which have been sequenced are trypsin inhibitors and naturally a major point of interest is to determine which residues are specifically involved in the inhibition, that is to define the $P_1$ site (1). As stated in the introduction the reason for sequencing the taro and giant swamp taro proteins was to see what differences in the inhibitor loop region might explain why they inhibit only trypsin, whereas the inhibitor from the giant taro is very effective against both trypsin and chymotrypsin. In our report on the sequence of the giant taro trypsin/chymotrypsin inhibitor (15), we indicated that in other Kunitz-type inhibitors from mono- and dicotyledons there is clear alignment of the $P_1$ sites whether for trypsin, chymotrypsin or subtilisin, with the amino acid residue varying accordingly. However for the aroid sequence such alignment has Glu61 in the equivalent position to the $P_1$ Arg63 residue of the soybean trypsin inhibitor (20). A glutamic acid residue is an unlikely site for trypsin or chymotrypsin inhibition and we noted that the aroid $P_1$ loop region is different from that of other Kunitz-type inhibitors. Following from a report of genetic engineering which selected a greatly improved neutrophil elastase inhibitor (21) we noted a striking similarity to their sequence of Ile-Ala-Phe-Phe-Pro, where the Ile is the $P_1$ site for the specific inhibition of neutrophil elastase. Residues 56-60 of the giant taro inhibitor are Leu-Ala-Phe-Phe-Pro and it was argued that with a Leu as the $P_1$ site the inhibitor might bind both trypsin and chymotrypsin and that this be the inhibitor site. The current sequencing shows that this region is conserved in all three aroid inhibitors and hence throws doubt on the choice of Leu56-Ala57 as the scissile bond. If Leu56 were the $P_1$ residue in the giant taro trypsin/chymotrypsin inhibitor, we need to postulate that there are sequence driven structural changes in the remaining loop region, which prevent the taro and giant swamp taro from having the broader specificity. Future work will concentrate on a chemical approach (20) to defining the $P_1$ sites in the aroid inhibitors.

## ACKNOWLEDGMENTS

We thank coworkers in the ACIAR/ANU Program on Root Crops in the Pacific for providing samples and the Australian Centre for International Agricultural Research for support. We also thank Dr Hirai for sending (on July 21, 1993) a preprint of his article, following his receipt from Dr P.J. Matthews (on March 12, 1993) of a copy of the information contained in our Fig. 1. Mr L. Peng is grateful for financial support from the International Foundation for Science, Sweden.

## REFERENCES

1. Laskowki, M.Jr. and Kato, I. (1980) Ann. Rev. Biochem. 50, 593-626.
2. Liener, I.E. and Kakade, M.L. (1980) in Toxic Constituents of Plant Foodstuffs (Liener, I.E., ed.) 2nd Ed. pp. 7-71, Academic Press, New York.
3. Plucknett, D.L. (1976) in Evolution of Crop Plants (Simmonds, N.W., ed.) p 10, Longman, London.
4. Chan, H.T. (ed.) (1983) Handbook of Tropical Foods, Preface, Marcel Dekker, New York.
5. Chandra, S. (ed.) (1984) Edible Aroids, Preface, Clarendon Press, Oxford.
6. Hammer, B.C. (1987) PhD. Thesis, Australian National University, Canberra, Australia.
7. Hammer, B.C., Shaw, D.C. and Bradbury, J.H. (1989) Phytochem. 28, 3019-3026.
8. Bradbury, J.H. and Hammer, B.C. (1990) J. Agric. Food Chem. 38, 1448-1453.
9. Sumathi, S. and Pattabiraman, T.N. (1975) Ind. J. Biochem. Biophys. 12, 383-385.
10. Sumathi, S. and Pattabiraman, T.N. (1979) Biochim. Biophys. Acta 566, 115-127.
11. Ogata, F. and Makisumi, S. (1984) J. Biochem. (Tokyo) 96, 1565-1574.
12. Ogata, F. and Makisumi, S. (1985) J. Biochem. (Tokyo) 97, 589-597.
13. Bradbury, J.H. and Holloway, W.D. (1988) Chemistry of Tropical Root Crops: Significance for Nutrition and Agriculture in the Pacific, 201 pp., Australian Centre for International Agricultural Research, Canberra, Australia.
14. Sumathi, S. and Pattabiraman, T.N. (1977) Biochim. Biophys. Acta 485, 167-178.
15. Argall, M.E., Bradbury, J.H., Hammer, B.C. and Shaw, D.C. (1993) Biochim. Biophys. Acta, submitted for publication.
16 Hirai, M., Nakamura, K., Imai, T. and Sato, T. (1993) Japanese J. Genetics, in press.
17 Needleman, S.B. and Wunsch, C.D (1970) J. Mol. Biol. 48, 443-453.
18 Gribskov, M. and Burgess, R.R. (1986) Nucl. Acids Res. 14(16), 6745-6763.
19 Plucknett, D.L. (1983) in Taro: A Review of *Colocasia esculenta* and its Potentials (Wang, J.K., ed.) p 14, University of Hawaii Press, Honolulu.
20 Ozawa, K. and Laskowski, M.Jr. (1966) J. Biol. Chem. 241, 3955-3961.
21 Roberts, B.L., Markland, W., Ley, A.C., Kent, R.B., White, D.W., Guterman, S.K. and Ladner, R.C. (1992) Proc. Natl. Acad. Sci. USA 89, 2429-2433.

*Phytochemistry*, Vol. 33, No. 1, pp. 21–27, 1993
Printed in Great Britain.

0031 9422/93 $6.00+0.00


# PURIFICATION AND PARTIAL CHARACTERIZATION OF TRYPSIN/CHYMOTRYPSIN INHIBITORS FROM CABBAGE FOLIAGE

ROXANNE M. BROADWAY

Department of Entomology, New York State Agricultural Experiment Station, Cornell University, Geneva, NY 14456, U.S.A.

(*Received in revised form* 25 *September* 1992)

**Key Word Index**—*Brassica oleracea*; Cruciferae; cabbage; protein characterization; proteinase inhibitors.

**Abstract**—Proteinase inhibitors from cabbage foliage (*Brassica oleracea*) had trypsin and chymotrypsin inhibitory activity that was relatively stable over a broad range of temperatures (0–100°) and pH values (4.5–7.5). The six proteinase inhibitors that were purified by affinity chromatography had $M_r$s that ranged from 9000 to 25000, and isoelectric points that ranged from 4.5 to 5.0. Separation of these affinity-purified proteins by reverse phase HPLC resulted in 14 unique protein species with trypsin and chymotrypsin inhibitory activity. Based on similarities in the amino acid content, the HPLC-purified inhibitors were arranged into four groups.

## INTRODUCTION

Serine proteinase inhibitor proteins are widely distributed in the plant kingdom [1], and inhibit the activity of enzymes (e.g. trypsin and chymotrypsin) that are commonly responsible for digestion in animals and microorganisms. In addition, these inhibitors have no apparent regulatory function in the plant. Thus, they are generally thought to contribute to the defence of plants against pathogens and herbivores [2, 3]. The most thoroughly examined plant serine proteinase inhibitors are found in species of Leguminosae, Gramineae and Solanaceae [1]. These inhibitors have been classified into at least seven non-homologous families based on their amino acid sequences [2, 4]. Proteinase inhibitory activity is present in the seeds, foliage and roots of cruciferous plants [5–7]. However, only a few seed proteins have been characterized [8–10]. Although the proteinase inhibitors in the foliage of crucifers have not been characterized, trypsin inhibitory activity in cabbage foliage has been shown to be developmentally regulated, appearing in high concentration in the young foliage on mature plants [11]. The present study describes the purification and partial characterization of the trypsin/chymotrypsin inhibitors in the young foliage of mature cabbage plants (*Brassica oleracea* L. cv Superpack).

## RESULTS

### *Purification of cabbage proteinase inhibitors*

After the initial extraction of proteins from young foliage on mature plants, affinity chromatography was used to separate the trypsin inhibitors from other contaminating proteins (Fig. 1), and HPLC was used to separate the trypsin inhibitors into at least 10 distinct peaks of protein that inhibited both trypsin and chymotrypsin (Fig. 2). There was a linear relationship between the *in vitro* concentration of inhibitor and the level of trypsin inhibited ($y=0.003+0.07x$, $r^2=0.982$).

Basing purification on increases in specific activity, a 67-fold purification of the proteins responsible for trypsin inhibitory activity was achieved following affinity chromatography (Table 1). With each purification step through affinity chromatography, the ratio of trypsin to chymotrypsin inhibitory activity remained the same (i.e. ratios were 1.7:1 for crude leaf juice; 1.9:1 for semi-purified PI; 1.8:1 for affinity-purified PI). This suggests that the chymotrypsin inhibitory activity originated from the same proteins that had affinity for trypsin, and therefore were 'double-headed' inhibitors, as has been reported for other plant proteinase inhibitors [1]. The dual activities were also noted for the inhibitor species separated by HPLC (Table 2), although the ratios of trypsin to chymotrypsin inhibitory activities differed among protein species. This difference in the ratios of trypsin to chymotrypsin inhibitory activities suggests there is a difference among the proteins in their binding capacity (i.e. number of binding sites, strength of binding) for trypsin and chymotrypsin.

### *Characterization of cabbage proteinase inhibitors*

*Stability of inhibitory activity.* Trypsin and chymotrypsin inhibitory activities in the semi-purified proteinase inhibitors were stable (100–71% activity) at high temperatures (100°) in a mildly acidic environment (1 mM HCl) for 10–120 min; and were significantly reduced (9–10% activity) only when autoclaved. However, the thermal stability of the trypsin inhibitory activity was significantly reduced when the inhibitor was incubated at 100°

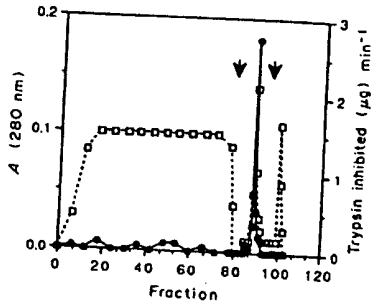


Fig. 1. Purification of cabbage proteinase inhibitors by affinity chromatography. The open squares indicate elution of protein, as measured by monitoring at 280 nm; the closed circles indicate the trypsin inhibitory activity in each fraction. The column was initially washed with 0.01 M Tris, pH 8.1, 0.1 M KCl; the proteinase inhibitors were eluted with 8 M urea, pH 3 (as indicated by the first arrow). Then the column was washed with 1 mM HCl, 0.1 M $CaCl_2$ (as indicated by the second arrow).

for 10 min in buffers outside the range of pH 4.5 to 7.5 (Fig. 3).

*Polyacrylamide gel electrophoresis.* Six protein bands were detectable in the green affinity-purified trypsin inhibitor on a non-denaturing, discontinuous polyacrylamide gel (Fig. 4), and these proteins all exhibited trypsin inhibitory activity (Fig. 5). The $M_r$s of these proteins, based on SDS–PAGE, were 25, 24, 20, 16, 13 and $12 \times 10^3$ (Fig. 6). The range of $M_r$ was confirmed by gel filtration (Fig. 7). The isoelectric points for the affinity-purified proteins were 4.5, 4.6, 4.7, 4.8, 4.9 and 5.0; four of these proteins banded as doublets (Fig. 8). The 'white affinity-purified trypsin inhibitor' only contained four proteins (Fig. 4) with $M_r$s of 21, 17, 15 and $11 \times 10^3$ (Fig. 6), and isoelectric points of 4.7, 4.8, 4.9 and 5.0 (three of these proteins banded as doublets; Fig. 8).

HPLC separation of the affinity-purified trypsin inhibitors resulted in 10 peaks of protein (Fig. 2). Polyacrylamide gel electrophoresis of the HPLC fractions indicated that four of these protein peaks contained two protein species each, while the remaining six peaks contained a single protein species, for a total of 14 protein species. However, only six protein bands were detectable in the affinity-purified trypsin inhibitors (Fig. 4). This discrepancy may have resulted from co-migration of

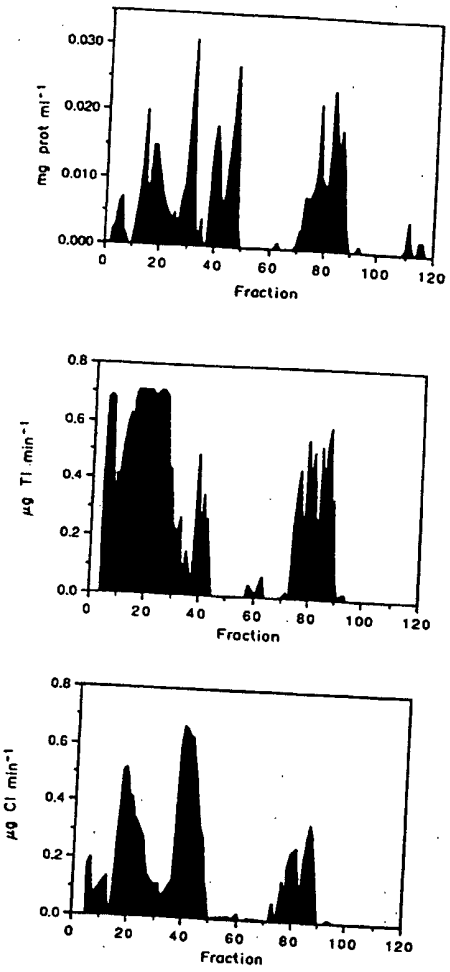


Fig. 2. HPLC separation of affinity-purified cabbage proteinase inhibitors. A 15-mg sample of affinity-purified proteinase inhibitor was applied to a preparative C-18 reverse-phase column (YMC ODS AQ 120, 30 × 250 mm), and eluted with a 22–52% linear acetonitrile gradient and 0.09% TFA (v/v); 10.5-ml fractions were collected. The three graphs, from top to bottom indicate: (a) the protein elution pattern; (b) elution pattern of trypsin inhibitory activity; and (c) elution pattern of chymotrypsin inhibitory activity. The units of activity for graphs b and c are μg of trypsin (or chymotrypsin) inhibited per min per fraction.

Table 1. Purification of trypsin inhibitors from 500 g of cabbage foliage

| Purification step | Total activity (mg trypsin inhibited) | Total protein (mg) | Specific activity | Recovery (%) | Purification (fold) |
|---|---|---|---|---|---|
| Leaf juice | 934 | 5130 | 0.18 | 100 | 1 |
| Semi-purified | 748 | 559 | 1.34 | 80 | 7 |
| Affinity-purified | 133 | 11 | 12.1 | 14 | 67 |

Table 2. Inhibitory activity and $M_r$ of HPLC-purified cabbage trypsin/chymotrypsin inhibitors

| HPLC fraction no. | mg tryp. inhib. (mg protein) | mg chymo. inhib. (mg protein) | Protein | $M_r \times 10^{-3}$ |
|---|---|---|---|---|
| 6 | 11 | 4 | Ia | 12 |
| | | | Ib | 9 |
| 14 | 3 | 0.5 | II | 10 |
| 18 | 6 | 4 | III | 10 |
| 31 | 0.2 | 0.5 | IV | 11 |
| 39 | 6 | 6 | Va | 12 |
| | | | Vb | 9 |
| 47 | 0 | 1 | VI | 10 |
| 74 | 5 | 0 | VII | 25 |
| 77 | 3 | 1 | VIIIa | 24 |
| | | | VIIIb | 11 |
| 83 | 1 | 0.4 | IXa | 24 |
| | | | IXb | 11 |
| 86 | 4 | 2 | X | 22 |

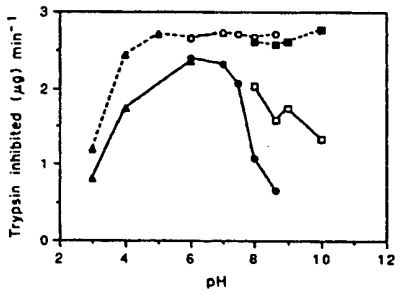


Fig. 3. Thermal and pH stability of cabbage proteinase inhibitors. A sample of semi-purified cabbage proteinase inhibitor (2 mg ml$^{-1}$ 1 mM HCl) was incubated in 1 ml of citrate phosphate (triangles), sodium phosphate (circles), or glycine–NaOH (squares) buffer, at 0° (dashed line) or 100° (solid line), for 10 min, then tested for trypsin inhibitory activity.

protein bands during gel electrophoresis of affinity-purified trypsin inhibitors, and/or biochemical modification of the proteins (e.g. glycosylation, de-amidation) that resulted in separate elution from the HPLC column. $M_r$s of the HPLC-purified protein species are indicated in Table 2.

*Amino acid analysis.* To ensure that the amino acid analyses for the immobilon-transferred proteins were as accurate as the proteins that were analysed directly, two protein species (fractions 18 and 77) were analysed by both methods. The results (data not shown) indicated that immobilon-transfer of protein did not introduce significant errors in the amino acid analysis. The amino acid analyses (Table 3) were organized into four groups based on similarities in amino acid content.

## DISCUSSION

Standard procedures (i.e. ammonium sulphate precipitation of aqueous extract, followed by affinity chromatography) were used to extract and purify six trypsin inhibitors from cabbage foliage. These inhibitors were relatively small ($M_r$ 9000–25 000), acidic (pI 4.5–5.0) proteins that were stable over a relatively broad range of pH values (4.5–7.5) and temperatures (0–100°). In general, these inhibitors had at least two active inhibitory sites: one specific for trypsin, the other specific for chymotrypsin. There was a difference between the number of trypsin inhibitors from non-chlorophyll-containing foliage (i.e. inner cabbage head) compared with chlorophyll-containing foliage. The inhibitors from the chlorophyll-containing foliage included two additional small, acidic proteins that were not present in the non-chlorophyll-containing foliage. At this point it is unclear whether the two additional proteins are found in green foliage, or are an experimental artifact resulting from proteolysis, or some other post-extractional biochemical modification. However, these proteins were consistently detected in each affinity-purified extract of young green foliage on mature plants, while they were absent from all extracts of white foliage.

The trypsin/chymotrypsin inhibitors from cabbage foliage are distinct from the serine proteinase inhibitors isolated from seeds of other crucifers, which are trypsin specific, basic proteins with unique amino acid contents [9, 12, 13].

The cabbage trypsin inhibitors were divided into four groups based on amino acid analyses (Table 3). The protein species in Group A were clustered, based on their large number of alanine and valine residues per mol of inhibitor, while the protein species in Group D were linked, based on the large number of arginine residues per mol of inhibitor. The four groups of inhibitors were also distinguished by the average per cent of hydrophobic amino acid residues: (A) $36 \pm 1.8\%$, (B) $31 \pm 1.1\%$, (C) $14 \pm 3.0\%$ and (D) $1.6 \pm 0.1\%$. However, amino acid sequences of these protein species are needed to determine the genetic relationship between these inhibitors.

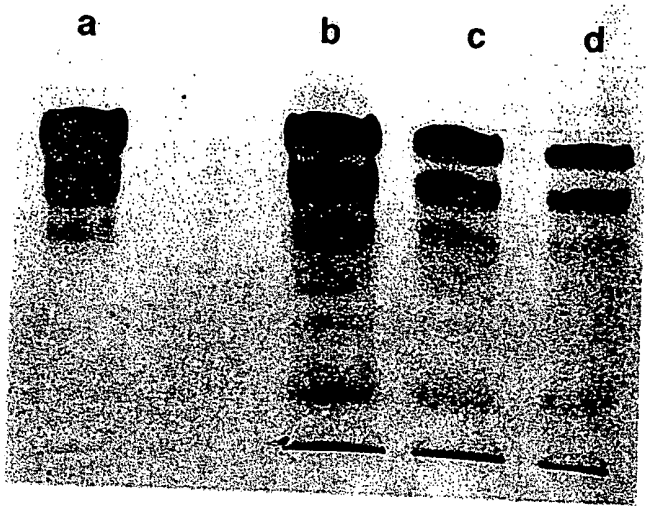


Fig. 4. Non-denaturing polyacrylamide gel electrophoresis of affinity-purified cabbage proteinase inhibitors. A discontinuous polyacrylamide gel (7.5% stacking gel, 20% separating gel) was used with non-denaturing buffer strips (0.88 M L-alanine, 0.25 M Tris, pH 8.8). Lane a contains white proteinase inhibitors, lanes b–d contain green proteinase inhibitor, at concentrations of 40, 20 and 10 $\mu$g per lane, respectively.

Fig. 5. Polyacrylamide gel electrophoresis of affinity-purified cabbage proteinase inhibitors. The gel consisted of 10% polyacrylamide, pH 8.8, with a 4% polyacrylamide stack, pH 6.8, 1.5 mm thick. The left portion of the gel was stained with Coomassie, while the right portion of the gel was incubated in a trypsin solution, then stained with acetyl-phenylalanine-$\beta$-naphthyl-ester, a trypsin-specific chromophoric substrate.

As in other plant families, such as Solanaceae and Leguminosae, the Cruciferae (e.g. cabbage) contained a large number of inhibitor species. Based on the fact that these inhibitors have differential binding capacity (i.e. inhibitory activity) for bovine trypsin and chymotrypsin (Table 2), perhaps this variability reflects a potential for inhibition of a number of different serine protease(s), from different species of herbivorous insects and/or plant pathogens.

## EXPERIMENTAL

*Plants.* Seeds for the cabbage cultivar Superpack (Johnny's Selected Seeds, Albion, ME) were germinated in Cornell Mix [14] in 7.6-l plastic pots. The seedlings were thinned to one plant per pot, and maintained in a greenhouse under 1 kW metal halide lamps (12 hr light: 12 hr dark) at 28°. The plants were watered three times a week, and fertilized once a week with a water-soluble nutrient mix (16N-32P-16K).

*Enzyme assays.* A standard spectrophotometric assay [15] was used to determine the presence of trypsin inhibitory activity in the cabbage extracts and purified fractions. Bovine trypsin (type III, Sigma) was 60% active as determined with the active site titrant, *p*-nitrophenyl-*p'*-guanidinobenzoate [16]. Bovine trypsin (0.1 mg ml$^{-1}$ 1 mM HCl) was mixed (1:1, v/v) with the plant proteinase inhibitor (2 mg ml$^{-1}$), and incubated at room temp. for 10 min. Then 100 $\mu$l of the mixt. was added to 2.9 ml of buffer (0.05 M Tris, pH 8.0) containing 1.04 M *p*-toluene–sulfonyl-L-arginine Me ester. Trypsin activity was monitored at 247 nm for 3 min, and compared with the activity of 50 $\mu$l of uninhibited trypsin. A micro-technique was used to measure trypsin inhibitory activity in affinity- and HPLC-purified proteins. Fractions were equilibrated based on protein content, then a 24-$\mu$l aliquot of the sample was mixed (1:1, v/v) with trypsin, as described above. The mixt. was added to 500 $\mu$l of substrate, and analysed for trypsin activity as described

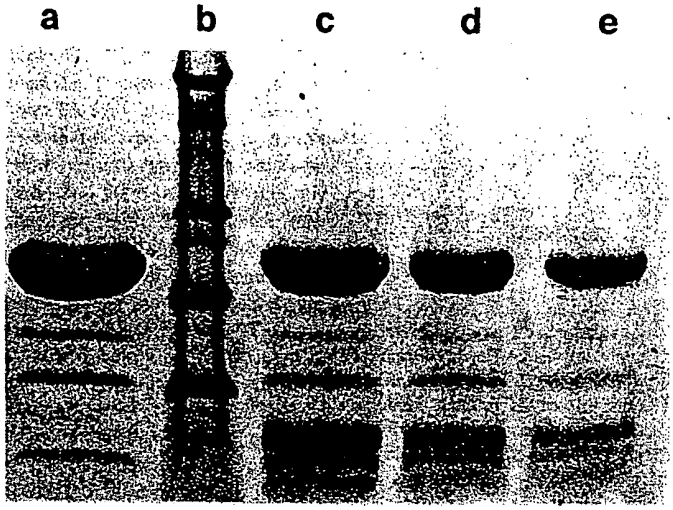


Fig. 6. SDS–PAGE of affinity-purified cabbage proteinase inhibitors. A discontinuous polyacrylamide gel (7.5% stacking gel, 20% separating gel) was used with buffer strips containing 0.2 M Tricine, 0.2 M Tris, 0.55% SDS, pH 8.1. Lane a contained white proteinase inhibitors, lane b contained $M_r$ markers, and lanes c–e contained green proteinase inhibitor, at concentrations of 40, 20 and 10 $\mu$g per lane, respectively. The $M_r$ markers were 66, 45, 36, 29, 24, 20.1 and 14.2 $\times 10^3$, while the proteinase inhibitors were determined to be 22–24, 17, 15, 12.5, 11 and 10.5 $\times 10^3$ for the green proteinase inhibitors, and 21–24, 17, 15 and 11 $\times 10^3$ for the white proteinase inhibitors.

Table 3. Amino acid analyses of HPLC-purified cabbage trypsin inhibitors

| | Group A | | | | Group B | | | | | Group C | | | Group D | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A′ | | A″ | | B′ | | B″ | | C′ | C″ | | C‴ | | |
| Protein | IXa | VII | VIIIa | X | II | III | Va | Ia | Vb | VIIIb | IXb | Ib | IV | VI |
| $M_r \times 10^{-3}$ | 24 | 25 | 24 | 22 | 10 | 10 | 12 | 12 | 9 | 11 | 11 | 9 | 11 | 10 |
| AA* | | | | | | | | | | | | | | |
| Asx | 17.5 | 18.7 | 18.6 | 18.0 | 9.8 | 9.6 | 12.9 | 10.8 | 6.7 | 6.1 | 7.5 | 4.7 | 5.4 | 5.4 |
| Glx | 22.6 | 21.4 | 21.3 | 20.0 | 8.0 | 8.2 | 9.5 | 10.7 | 9.2 | 10.6 | 11.8 | 6.9 | 12.3 | 11.6 |
| Ser† | 22.0 | 16.9 | 16.1 | 14.3 | 2.9 | 2.9 | 5.0 | 9.8 | 5.2 | 3.7 | 2.9 | 3.1 | 0.2 | 0.1 |
| Gly | 44.8 | 43.5 | 27.8 | 26.1 | 11.5 | 12.1 | 15.2 | 16.1 | 15.8 | 19.0 | 18.9 | 51.5 | 19.8 | 17.6 |
| His | 1.6 | 0.7 | 0.3 | 0.1 | 1.0 | 0.8 | 1.5 | 0.7 | 0.9 | 0.7 | 0.5 | 0.5 | 0.1 | 0.1 |
| Lys | 12.2 | 12.5 | 12.7 | 12.9 | 6.5 | 7.8 | 7.9 | 7.6 | 5.0 | 5.5 | 6.2 | 4.2 | 6.0 | 5.4 |
| Arg | 14.5 | 13.7 | 14.1 | 13.6 | 6.1 | 7.6 | 6.7 | 9.3 | 13.3 | 21.7 | 24.3 | 12.7 | 31.7 | 28.4 |
| Thr† | 11.3 | 11.4 | 12.1 | 11.1 | 3.6 | 4.2 | 5.2 | 3.0 | 5.6 | 8.7 | 9.6 | 4.8 | 12.8 | 11.5 |
| Cys† | 1.0 | 0.0 | 1.2 | 3.0 | 8.5 | 7.4 | 4.1 | 4.2 | 0.0 | 0.0 | 0.9 | 0.9 | 5.0 | 1.2 |
| Ala | 13.1 | 12.1 | 12.2 | 10.6 | 2.7 | 2.9 | 4.4 | 4.0 | 2.5 | 1.8 | 1.0 | 1.3 | 0.2 | 0.1 |
| Pro | 12.7 | 15.2 | 16.5 | 17.1 | 7.0 | 6.1 | 7.1 | 4.2 | 1.6 | 2.0 | 0.7 | 1.0 | 0.1 | 0.1 |
| Tyr | 8.4 | 12.5 | 9.9 | 5.2 | 3.8 | 4.1 | 6.1 | 8.5 | 5.1 | 3.9 | 2.8 | 4.2 | 0.0 | 0.0 |
| Val | 13.7 | 17.0 | 17.5 | 16.8 | 5.4 | 5.5 | 5.1 | 3.5 | 2.0 | 2.3 | 1.1 | 1.0 | 0.2 | 0.2 |
| Met† | 0.3 | 0.0 | 0.1 | 0.9 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.6 | 0.5 |
| Ile | 9.6 | 11.3 | 11.3 | 9.9 | 2.7 | 2.5 | 2.5 | 3.5 | 1.8 | 1.7 | 1.2 | 1.1 | 0.2 | 0.1 |
| Leu | 14.5 | 14.5 | 14.3 | 13.0 | 4.1 | 4.1 | 6.2 | 4.8 | 3.7 | 4.1 | 2.2 | 1.7 | 0.2 | 0.1 |
| Phe | 11.5 | 14.8 | 15.2 | 14.1 | 5.3 | 5.9 | 6.8 | 4.8 | 2.1 | 1.0 | 0.0 | 0.6 | 0.2 | 0.2 |

*Amino acid content expressed as residues $mol^{-1}$ of inhibitor.
†Acid hydrolysis is destructive to these amino acids.

above. A 12-$\mu$l aliquot of trypsin was used to measure uninhibited trypsin activity.

Chymotrypsin inhibitory activity was determined by mixing the plant extract or purified sample (at the same concentrations indicated above) 1:1 (v/v), with TLCK-treated bovine chymotrypsin (0.1 mg $ml^{-1}$ 1 mM HCl) for 10 min at room temp. Then, 80 $\mu$l of the mixt. was added to 2.9 ml of substrate [1 mM benzoyl-L-tyrosine Et ester in 5% MeOH, mixed 1:1 (v/v) with 0.05 M Tris, pH 8], and monitored at 256 nm for 3 min [17]. The results

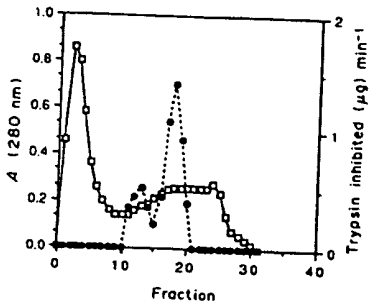


Fig. 7. Estimation of $M_r$s of cabbage proteinase inhibitors by liquid chromatography on Sephadex G-75. The open squares indicate protein elution; the closed circles indicate trypsin inhibitory activity. The $M_r$ markers eluted as follows: carbonic anhydrase ($29 \times 10^3$) in fraction 10, soybean trypsin inhibitor ($20.1 \times 10^3$) in fraction 12 and cytochrome c ($12.4 \times 10^3$) in fraction 16.

were compared with the activity of 40 μl of uninhibited chymotrypsin. The chymotrypsin inhibitory activity in affinity- and HPLC-purified proteins was measured by the micro-technique described for measuring trypsin inhibitory activity.

*Protein determination.* Protein concn was estimated using bicinchoninic acid reagent [18]. Purified protein from cabbage foliage was used as the standard.

### *Purification of proteinase inhibitors*

*Extraction.* Proteinase inhibitors were extracted from mature cabbage plants, using either fresh, young foliage (including the cabbage head and surrounding green leaves) or only the white foliage of the cabbage head. A single extraction consisted of homogenization of 500 g of foliage in 500 ml of ice-cold, 0.1 M Tris–HCl buffer, pH 7, 0.1% ascorbic acid (w/v). The homogenate was squeezed through a double layer of cheesecloth, and the liquid was centrifuged at 4200 *g* for 10 min at 4°. $(NH_4)_2SO_4$ was added to the supernatant at 80% satn, and incubated at 4° for 1 hr with stirring. After incubation, the mixt. was centrifuged at 6000 *g*, 4°, for 20 min, then the pellet was resuspended in $H_2O$, and dialysed ($M_r$ cut-off 12 000–14 000) against $H_2O$ at 4°. The dialysate was centrifuged, and the supernatant was lyophilized. This preparation was called 'semi-purified proteinase inhibitors'. The material that originated from total young foliage on mature plants was designated 'green semi-purified proteinase inhibitors', and the material from the white foliage in the cabbage head was designated 'white semi-purified proteinase inhibitors'.

*Affinity chromatography.* Semi-purified proteinase inhibitor (100 mg 50 $ml^{-1}$ 0.01 M Tris, pH 8.1, 0.1 M KCl) was applied to an affinity column (bovine trypsin, bound to cyanogen bromide-activated Sepharose 4B, 1.5 × 33 cm) at 4°. The column was washed with 0.01 M Tris, pH 8.1, 0.1 M KCl, until the *A* at 280 nm approached zero. Then, the proteinase inhibitor was eluted with 8 M urea, pH 3. The fractions from the entire protein peak (*A* at 280 nm) were pooled, dialysed against $H_2O$ at 4°, then lyophilized.

### *Characterization of cabbage proteinase inhibitors*

*Polyacrylamide gel electrophoresis.* The PhastSystem electrophoresis unit (Pharmacia) was used to determine the number of protein bands, $M_r$ and isoelectric point of the affinity-purified trypsin inhibitors. The PhastSystem was also used to determine the $M_r$ (20% discontinuous polyacrylamide PhastGel, with Tris/SDS buffer strips) of

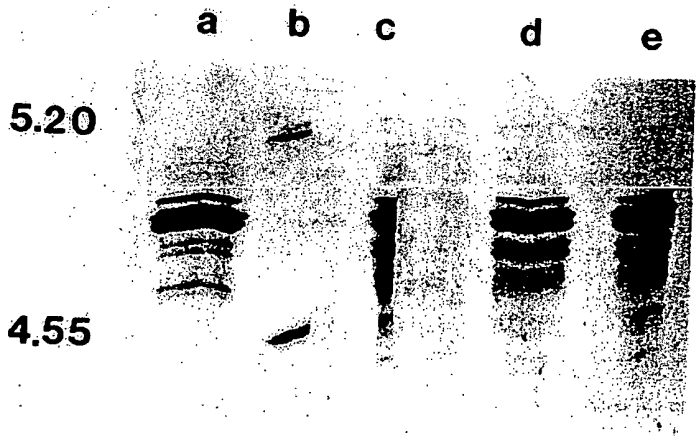


Fig. 8. Isoelectric focusing of affinity-purified proteinase inhibitors. A 5% polyacrylamide gel containing ampholytes for pI 4–6.5 was used. The isoelectric standards were pI 3.5, 4.55, 5.2, 5.85, 6.55, 6.85, 7.35, 8.15, 8.45, 8.65 and 9.3. Lane a contains white proteinase inhibitors, lane b contains isoelectric focusing markers and lanes c–e contain green proteinase inhibitor at concentrations of 40, 20 and 10 μg per lane, respectively. The isoelectric points for the green affinity-purified proteinase inhibitors were 4.5, 4.6, 4.7, 4.8, 4.9 and 5.0 while the white affinity-purified proteinase inhibitors were 4.7, 4.8, 4.9 and 5.0.

HPLC-purified trypsin inhibitors (a 5-$\mu$g aliquot of each protein). Each test sample was dried in a Speed-Vac, resuspended in 3 $\mu$l deionized, distilled $H_2O$ + 1 $\mu$l 4 × sample buffer [40 mM Tris, 4 mM EDTA, pH 8.0, 20% 2-mercaptoethanol, 0.04% bromophenyl blue (w/v)], then placed in boiling water for 5 min prior to SDS–PAGE analysis.

To determine the purity of the affinity-purified trypsin inhibitor, a 75-$\mu$l aliquot of a 4 mg ml$^{-1}$ soln of affinity purified trypsin inhibitor was applied to a vertical, discontinuous, non-denaturing, 1.5 mm, 10% polyacrylamide gel, pH 8.8, with a 4% polyacrylamide stacking gel, pH 6.8. The reservoir buffer was 0.025 M Tris, pH 8.3, 0.192 M glycine. Following electrophoresis, the gel was stained with Coomassie Brilliant Blue R (20% MeOH, 30% HOAc, 0.1% Coomassie), and destained with 30% MeOH, 10% HOAc to detect all protein bands. To detect protein bands with trypsin inhibitory activity, a modification of the method from ref. [19] was used. The polyacrylamide gel was washed four times with 30% MeOH, 20% HOAc to fix the protein, then rinsed with $H_2O$ twice. The gel was equilibrated overnight in 0.1 M NaPi buffer, pH 8, then incubated in a trypsin soln (0.1 mg trypsin ml$^{-1}$ 0.1 M NaPi buffer, pH 8) for 30 min at 37°. The gel was rinsed twice with $H_2O$, then covered with a freshly prepared soln of 2.5 mg acetyl-phenylalanine-$\beta$-naphthyl-ester (APNE) in 1 ml dimethylformamide + 9 ml 0.55 mg ml$^{-1}$ tetrazotized *o*-dianisidine, in 0.1 M NaPi, pH 8. The gel was incubated in the APNE soln for 30 min at 37°, then rinsed with $H_2O$. Digestion of APNE by trypsin resulted in the release of a red chromophore. Clear bands indicated the inhibition of trypsin. A lane containing BSA (no trypsin inhibitory activity) was run as a control.

*Amino acid analysis.* The amino acid analysis of each HPLC-purified protein was performed at the Cornell Amino Acid Facility on the Waters Pico-Tag HPLC System [20]. HPLC fractions containing a single protein (as determined by SDS–PAGE) were analysed for amino acid composition directly as follows. A 10-$\mu$g aliquot of each fraction was dried in a Speed-Vac, acid hydrolysed (6 M HCl, 95 min, 150°, under partial vacuum and $N_2$), then analysed for amino acid composition. HPLC fractions containing more than one protein band were electrophoresed, to separate the proteins prior to amino acid analysis. A 10-$\mu$g aliquot of each fraction was dried in the Speed-Vac, electrophoresed on SDS–PAGE, as described above (PhastSystem, Pharmacia), then the proteins were electrophoretically transferred to PVDF transfer membrane (Immobilon-P, 0.45 $\mu$m, Millipore) using the PhastSystem Transfer Unit and Tris–glycine transfer buffer (25 mM Tris, 192 mM glycine, pH 8.5, 20% MeOH). The Immobilon was stained with Coomassie (0.1% Coomassie, 60% MeOH, 7.5% HOAc), destained with 75% MeOH, 7.5% HOAc, then exhaustively rinsed with deionized, distilled $H_2O$, to remove excess glycine originating from the transfer buffer. Each protein band was acid hydrolysed (6 M HCl, 115 min, 150°, under partial vacuum and $N_2$), the amino acids were extracted from the Immobilon, then analysed for amino acid composition [20].

*Acknowledgements*—This research was supported, in part, by Cornell Biotechnology Program (BRO 1), which is sponsored by the New York State Science and Technology Foundation; U. S. Department of Agriculture (91-00843); and by HATCH funds. I wish to thank A. A. Colvin and W. C. Kain for outstanding technical support, and Drs S. S. Duffey and David M. Soderlund for critical review of this manuscript.

## REFERENCES

1. Whitaker, J. R. (1981) in *Impact of Toxicology on Food Processing* (Ayres, J. C. and Kirschman, J. C., eds), p. 57. AVI Publishing, Westport, CT.
2. Ryan, C. A. (1989) *BioEssays* **10**, 20.
3. Geoffroy, P., Legrand, M. and Fritig, B. (1990) *Molec. Plant–Microbe Interact.* **3**, 327.
4. Laskowski, M. and Kato, I. (1980) *A. Rev. Biochem.* **49**, 593.
5. Chen, I. and Mitchell, H. L. (1973) *Phytochemistry* **12**, 327.
6. Broadway, R. M. (1989) *Phytochemistry* **28**, 755.
7. Hojima, Y., Pisano, J. J. and Cochrane, C. G. (1983) *Biochem. Pharmac.* **32**, 985.
8. Carter, T. H., Everson, B. A. and Ratnoff, O. D. (1990) *Blood* **75**, 108.
9. Wilimowska-Pelc, A. (1985) *Acta Biochim. Polon.* **32**, 351.
10. Ogawa, T., Higasa, T. and Hata, T. (1968) *Agric. Biol. Chem.* **32**, 484.
11. Broadway, R. M. and Missurelli, D. L. (1990) *Phytochemistry* **29**, 3721.
12. Ogawa, T., Higasa, T. and Hata, T. (1971) *Agric. Biol. Chem.* **35**, 712.
13. Menegatti, E., Palmieri, S., Walde, P. and Luisi, P. L. (1985) *J. Agric. Food Chem.* **33**, 784.
14. Boodley, J. W. and Sheldrake Jr, R. (1977) *Cornell Information Bull.* **43**, 1.
15. Walsh, K. A. (1970) *Meth. Enzym.* **19**, 41.
16. Chase, T. J. and Shaw, E. (1967) *Biochem. Biophys. Res. Commun.* **29**, 508.
17. Walsh, K. A. and Wilcox, P. E. (1970) *Meth. Enzym.* **19**, 31.
18. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J. and Klenk, D. C. (1985) *Analyt. Biochem.* **150**, 76.
19. Filho, J. X. and Moreira, R. A. (1978) *Analyt. Biochem.* **84**, 296.
20. Cohen, S. A., Tarvin, T. L., Bidlingmeyer, B. A. and Tarr, G. E. (1984) *Am. Lab.* Aug., 48.